

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model
Run on: July 13, 2004, 12:02:13 ; Search time 88 Seconds
(without alignments)
12247.843 Million cell updates/sec

Title: US-09-270-437D-5
Perfect score: 3110
Sequence: 1 aggcagcgcgcgcaccgccc.....attctcttcaggttttaaaa 1708

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1017041 segs, 315518202 residues
Total number of hits satisfying chosen parameters: 2034082

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+n2p.model -DEV=xlp
-Q=/cgn2_1/USFTO_spool_p/US09270437/runat_13072004_121921_9604/app_query.fasta_1.1863
-DB=SPREMBL_25 -QFMT=fastan -SUFFIX=rspt -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -NATX=D-BLOSUM62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORX=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09270437@cgn 1.1 146 @runat_13072004_121921_9604 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

- Database :
- 1: sp_archaea:*
 - 2: sp_bacteria:*
 - 3: sp_fungi:*
 - 4: sp_human:*
 - 5: sp_invertebrate:*
 - 6: sp_mammal:*
 - 7: sp_mhc:*
 - 8: sp_organelle:*
 - 9: sp_phage:*
 - 10: sp_plant:*
 - 11: sp_rodent:*
 - 12: sp_virus:*
 - 13: sp_vertebrate:*
 - 14: sp_unclassified:*
 - 15: sp_rvirus:*
 - 16: sp_bacteriap:*
 - 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Length	DB ID	Description
1	2221	71.4	577 11 Q8CGX0	Q8cgx0 rattus norv

2	2216	71.3	577 4	Q9NZ18
3	2208	71.0	577 11	O88477
4	2201	70.8	577 11	O8BRH1
5	2201	70.8	577 11	O80US9
6	2116.5	68.1	576 13	O42254
7	1716.5	55.2	594 13	O73932
8	1716	55.2	593 13	O57526
9	1637	52.6	579 4	O00425
10	1628	52.3	579 11	Q9CPN8
11	1621	52.1	579 11	Q8CJ98
12	1612.5	51.8	582 13	Q9PW80
13	1464	47.1	556 4	Q9Y6M1
14	1073	34.5	545 11	Q7TQF9
15	858.5	27.6	566 5	Q9VZ69
16	858	27.6	573 5	Q8IR99
17	853.5	27.4	580 5	Q8IGK4
18	459.5	14.8	828 5	Q21605
19	257.5	8.3	621 10	Q9C553
20	257.5	8.2	398 3	Q7TF50
21	256.5	8.1	454 10	Q84MA6
22	251	8.1	479 10	Q9XI71
23	251	8.1	479 10	Q9XI71
24	248	8.0	557 5	Q23487
25	246.5	7.9	510 5	Q7Z145
26	242	7.8	318 13	Q7ZVX5
27	240.5	7.7	680 5	P91393
28	240	7.7	672 13	Q8AX85
29	238	7.7	542 10	Q8S7G1
30	237.5	7.6	762 10	Q8LNT9
31	236.5	7.6	641 5	Q9BLA0
32	235.5	7.6	632 10	Q82762
33	234.5	7.5	568 10	Q9LXF5
34	233.5	7.5	313 4	Q96EP6
35	233	7.5	653 13	Q7ZXS1
36	232	7.5	589 5	Q17935
37	232	7.5	611 5	Q17936
38	231	7.4	640 10	Q9ASX3
39	231	7.4	644 10	Q9PNK3
40	230	7.4	774 10	Q9L128
41	228	7.3	644 10	Q8LDV1
42	226.5	7.3	577 10	Q9SR13
43	224	7.2	370 11	Q8BSB0
44	222.5	7.2	364 5	Q95SZ9
45	220.5	7.1	353 13	Q9W702

ALIGNMENTS

RESULT 1

ID	Q8CGX0	PRELIMINARY;	PRT;	577 AA.
AC	Q8CGX0;			
DT	01-MAR-2003 (TrEMBLrel. 23, Created)			
DT	01-MAR-2003 (TrEMBLrel. 23, Last sequence update)			
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)			
DE	B-actin zipcode binding protein 1.			
OS	Rattus norvegicus (Rat)			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
OX	NCBI_TaxID=10116;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=Sprague-Dawley;			
RA	Eom T., Singer R.H., Bassell G.J.;			
RT	"Molecular interactions between r2Ep1 and b-actin zipcode required for transport of mRNA and stimulation of spine growth.";			
RL	Submitted (AUG-2002) to the EMBL/GenBank/DDeU databases.			
DR	EMBL; AF541940; AAC16210.1;			
DR	GO; GO:0003676; F:nucleic acid binding; IEA.			
DR	InterPro; IPR004087; KH dom.			
DR	InterPro; IPR004088; KH type 1.			
DR	InterPro; IPR000504; RNA_rec_mot.			
DR	Pfam; PF00013; KH; 4.			

DR Pfam; PF00076; rtm; 2.
 DR SMART; SM00322; KH; 4.
 DR SMART; SM00360; RRM; 2.
 DR PROSITE; PS50084; KH TYPE_1; 4.
 DR PROSITE; PS50102; RRM; 2.
 SQ SEQUENCE 577 AA; 63436 MW; 0647676128FBD1EE CRC64;

Alignment Scores:

Pred. No.: 3, 81e-160 Length: 577
 Score: 2221.00 Matches: 440
 Percent Similarity: 99.10% Conservatism: 1
 Best Local Similarity: 98.88% Mismatches: 2
 Query Match: 71.41% Indels: 2
 DB: 11 Gaps: 1

US-09-270-437D-5 (1-1708) x Q8CGX0 (1-577)

QY 32 CGGGAGCCATCATGAAGCTGAATGGCCACAGTTGGAGAACCATGCCCTGAAGGTCTCC 91
 DB 133 ArgGlnAlaIleMetLysLeuAsnGlyHisGlnLeuGluAsnHisAlaLeuLysValSer 152
 QY 92 TACATCCCGATGACAGATAGACAGCGGACCTGAGATGGCGCGCGGGGCTTTGGC 151
 DB 153 TyrIleProAspGluGlnIleAlaGlnGlyProGluAsnGlyArgArgGlyGlyPheGly 172
 QY 152 TCTCGGGTCAGCCGCCCGCCAGGGCTCACCTGTGGCAGCGGGGCCCGCCAGCAGCAG 211
 DB 173 SerArgGlyGlnProArgGlnGlySerProValAlaAlaGlyAlaProAlaLysGlnGln 192
 QY 212 CAGTGGACATCCCGTCCCTGGCTCCCTGGTCCCGCCAGTATGTGGTGCCTATTATGGC 271
 DB 193 GlnValAspIleProLeuArgLeuValProThrGlnTyrValGlyAlaIleLeGly 212
 QY 272 AAGGAGGGGCCACCATCCGCAACATCACAAAAACAGACCCAGTCCCAAGATAGACGTGCAT 331
 DB 213 LysGluGlyAlaThrIleArgAsnIleThrLysGlnThrGlnSerLysIleAspValHis 232
 QY 332 AGGAGGAGACAGCAGGTGAGTGAAGAAAGCCATCAGTGTGCTCCACCCCTGAGGGC 391
 DB 233 ArgLysGluAsnAlaGlyAlaAlaGluLysAlaIleSerValHisSerThrProGluGly 252
 QY 392 TGCTCTCCCTTGTAAGATGATCTTGGAGATTATGCATAAAGAGGCTAAGGACACCCAA 451
 DB 253 CysSerSerAlaCysLysMetIleLeuGluIleMetHisLysGluAlaLysAspThrLys 272
 QY 452 ACGGCTGACAGGTTCCCTGAAGATCTTGGGCCATATAACTTTGTAGGGCGTCTCAT 511
 DB 273 ThrAlaAspGluValProLeuLysIleLeuAlaHisAsnAsnPheValGlyArgLeuIle 292
 QY 512 GGCAGGAGCAGGACCTGAGAGGTAGAGCAGATACCGACACAAAATACCATC 571
 DB 293 GlyLysGluGlyArgAsnLeuLysValGluGlnAspThrGluThrLysIleThrIle 312
 QY 572 TCTCTGTTGCAAGACTTACCTTTTACAAACCTGAGAGGACCATCACTGTGAAGGGGGCC 631
 DB 313 SerSerLeuGlnAspLeuThrLeuTyrAsnProGluArgThrIleThrValLysGlyAla 332
 QY 632 ATCGAGAAATGTGCGGGCCGCGAGCGGAAATATGAGAAAGTTCCGGAGGCGCTATGAG 691
 DB 333 IleGluAsnCysCysArgAlaGluGlnGluIleMetLysLysValArgGluAlaTyrGlu 352
 QY 692 AATGATGTGGCTGCCATGAGC-----TCTCACTGATCCCTGGCTGCAACTGGCTGCT 745
 DB 353 AsnAspValAlaAlaMetSerLeuGlnSerHisLeuIleProGlyLeuAsnLeuAlaAla 372
 QY 746 GTAGGTCTTTCCAGCTTCATCCAGCGAGTCCCGCGCTCCCGAGCGGCTTACTGGG 805
 DB 373 ValGlyLeuPheProAlaSerSerAlaValProProProSerSerValThrGly 392
 QY 806 GCTGTCTCCATAGCTCTCTTTATGCAAGGCTCCCGAGCAGGAGATGGTGCAGGTGTTTATC 865
 DB 393 AlaAlaProTyrGlySerPheMetGlnAlaProGluGlnGluMetValGlnValPheIle 412

QY 866 CCGGCCAGGCGAGTGGGGCCCATCATCGGCAAGAGGGCGAGCAGCATCAAAACAGCTTCC 925
 DB 413 ProAlaGlnAlaValGlyAlaIleIleGlyLysLysGlyGlnHisIleLysGlnLeuSer 432
 QY 926 CGGTTTGGCAGCGCTCCATCAAGATTGACACCCAGCAACACCTGACTCCCAAGTTTGGT 985
 DB 433 ArgPheAlaSerAlaSerIleLysIleAlaProProGluThrProAspSerLysValArg 452
 QY 986 ATGTTATCATCATCTGGACCCCGAGCCCAATTCAGGCTCAGGGAAGAATCTATGGC 1045
 DB 453 MetValValIleThrGlyProProGluAlaGlnPheLysAlaGlnGlyArgIleTyrGly 472
 QY 1046 AAACCTCAGAGGAGAGACTCTTTGGTCCCGAGGAGGAGTGAAGCTGGAGACCCACATA 1105
 DB 473 LysLeuLysGluGluAsnPheGlyProLysGluGluValLysLeuGluThrHisIle 492
 QY 1106 CGTGTGCCAGCATCAGCAGCTGGCGGGCTCATTTGCCAAAGGTGAAAAACGGTGAAACGAG 1165
 DB 493 ArgValProAlaSerAlaAlaGlyArgValIleGlyLysGlyLysThrValAsnGlu 512
 QY 1166 TTGCAGAAATTTGACGGCAGCTGAGTGGTAGTACCAAGAGACACAGCCCTGATGAGAAC 1225
 DB 513 LeuGlnAsnLeuThrAlaAlaGluValValProArgAspGlnThrProAspGluAsn 532
 QY 1226 GACCAGGTTCATCGTGAATAATCATCGGACATTTCTATGCCAGTCAAGTGGCTCAACGGAAG 1285
 DB 533 AspGlnValIleValLysIleIleGlyHisPheTyrAlaSerGlnMetAlaGlnArgLys 552
 QY 1286 ATCGAGACATCTCGGCCCGAGTTAAGCAGCAGCATCAGAGGAGACAGATAACCGAGCC 1345
 DB 553 IleArgAspIleLeuAlaGlnValLysGlnHisGlnLysGlyGlnSerAsnGlnAla 572
 QY 1346 CAGCAGCGGAGGAG 1360
 DB 573 GlnAlaArgArgLys 577

RESULT 2

Q9NZ18 PRELIMINARY; PRT; 577 AA.
 AC Q9NZ18;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE mRNA-binding protein CRDBP.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Ioannidis P., Tringas T., Dimitriadis E., Samiotaki M.,
 RA Panoutsakopoulos G., Kyriazoglou I., Voutzoulis S., Tsiapalis C.M.,
 RA Kittas C., Agnantis N., Pandis N.;
 RT "Ectopic expression of a KH-domain containing protein, highly
 RT homologous to both human IMP-1 and mouse CRD-BP, in benign and
 RT malignant mesenchymal tumors.";
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
 CC -!- SIMILARITY: CONTAINS 4 KH DOMAINS.
 DR EMBL; AF198254; AAF37203.1; -;
 DR HSP; P11940; 1CVJ.
 DR GO; GO:0003576; F: nucleic acid binding; IEA.
 DR InterPro; IPR004087; KH_dom.
 DR InterPro; IPR004086; KH_type_1.
 DR InterPro; IPR000504; RNA_rec_mot.
 DR Pfam; PF00013; KH; 4.
 DR Pfam; PF00076; rtm; 2.
 DR SMART; SMC0322; KH; 4.
 DR SMART; SMC0360; RRM; 2.
 DR PROSITE; PS50084; KH TYPE_1; 4.
 DR PROSITE; PS50102; RRM; 2.
 SQ SEQUENCE 577 AA; 63456 MW; 0749A060F252D81D CRC64;

Alignment Scores:

Pred. No.: 9,15e-160 Length: 577
 Score: 2216.00 Matches: 440
 Percent Similarity: 98.88% Conservative: 0
 Best Local Similarity: 98.88% Mismatches: 3
 Query Match: 71.25% Indels: 2
 DB: 4 Gaps: 1

US-09-270-437D-5 (1-1708) x Q9NZ18 (1-577)

QY 32 CGGGAGCCATCATGAAGCTGAATGCGCCACCAAGTGGAGAACCATGCCCTCGAAGTCTCC 91
 Db 133 ArgGlnAlaIleMetLysLeuAsnGlyHisGlnLeuGluAsnHisAlaLeuLysValSer 152
 QY 92 TACATCCCGATGAGCAGATAGACAGGACCTGAGATGGCGCGCGAGGGGCTTGGC 151
 Db 153 TyrIleProAspGluGlnIleAlaGlnGlyProGluAsnGlyArgargGlyGlyPheGly 172
 QY 152 TCTCGGGGTGAGCCCGCGAGGCTCAGCTGTGGCAGCGGGGGCCCGCCAGCCAGCAGCAG 211
 Db 173 SerArgGlyGlnProArgGlnGlySerProValAlaAlaGlyAlaProAlaLysGlnGln 192
 QY 212 CAAGTGGACATCCCTCGCTGCTGGTCCGCCACCCAGTATGTGGTCCCATATTGCG 271
 Db 193 GlnValAspIleProLeuArgLeuLeuValProThrGlnTyrValGlyAlaIleGly 212
 QY 272 ARGAGGGGCGCCACCATCCGCAACATCACAAACAGACCCAGTCCCAAGATAGACGTGCAT 331
 Db 213 LysGluGlyAlaThrIleArgAsnIleThrLysGlnThrGlnSerLysIleAspValHis 232
 QY 332 AGGAAGGAGAACCGAGTGCAGCTGAGTGAAGAACCCATCAGTGTGCATCCACCCCTGAGGCG 391
 Db 233 ArgLysGluAsnAlaGlyAlaAlaGluLysAlaIleSerValHisSerThrProGluGly 252
 QY 392 TGCTCTCCCTGTTAAGATGATCTGGAGATTATGCATAAGACAGCTAAGGACACCAAA 451
 Db 253 CysSerSerAlaCysLysMetIleLeuGluIleMetLysLysGluAlaLysAspThrLys 272
 QY 452 ACGGCTCAGAGGTTCCCTCGAAGATCTCGGCCCATTAATAACTTGTAGGGCTCTCAT 511
 Db 273 ThrAlaaspGluValProLeuLysThrLeuAlaHisAsnAsnPheValGlyArgLeuIle 292
 QY 512 GCGAAGGAGACGGAACCTGAAGAGGTAGAGAGTACAGACATCCAGACAAATATCACAATC 571
 Db 293 GlyLysGluGlyArgAsnLeuLysLysValGluGlnAspThrGluThrLysIleThrIle 312
 QY 572 TCCTCGTTGCAAGACCTTACCTTTACACCCCTGAGAGGACCATCACTGTGAAGGGGGCC 631
 Db 313 SerSerLeuGlnAspLeuThrLeuTyrAsnProGluArgThrIleThrValLysGlyAla 332
 QY 632 ATCAGAGATTGTTTCAGGGCGCGACGAGAAATATGAAGAAAGTTCGGAGGCGCTATGAG 691
 Db 333 IleGluAsnCysCysArgAlaGluGlnGluIleMetLysLysValArgGluAlaTyrGlu 352
 QY 692 AATGATGGTGCATGAGC-----TCTCAGCTGATCCCTGGCCCTGACCTGGCTGCT 745
 Db 353 AsnAspValAlaAlaMetSerLeuGlnSerHisLeuThrProGlyLeuAsnLeuAlaAla 372
 QY 746 GTAGGCTTTTCCAGGTTTCATCCAGGCGAGTCCCGCGGCTCCAGCAGCGTACTGGG 805
 Db 373 ValGlyLeuPheProAlaSerSerSerAlaValProProProProSerValThrGly 392
 QY 806 GCTGCTCCCTATAGCTCCTTATGAGGCTCCCGAGCAGAGATGCTGCGAGGTGTTATC 865
 Db 393 AlaAlaProTyrSerSerPheMetGlnAlaProGluGlnGluMetValGlnValPheIle 412
 QY 866 CCGCGCCAGGCGAGTGGGGCCCATCATCGGCAAGAGGGGCGAGCACATCAACAGCTCTCC 925
 Db 413 ProAlaGlnAlaValGlyAlaIleIleGlyLysLysGlyGlnHisIleLysGlnLeuSer 432
 QY 926 CGGTTTGGCCGCTCCATCAAGATTGACACCCGAAACACCTGACTGCTCAAGATTCGT 985
 Db 433 ArgPheAlaSerAlaSerIleLysIleAlaProProGluThrProAspSerLysValArg 452

QY 986 ATGGTTATCATCTGCGACCGCCAGAGGCCCAATTCAAGGCTCAGGGAAGAATCTATGGC 1045
 Db 453 MetValIleIleThrGlyProProGluAlaGlnPheLysAlaGlnGlyArgIleTyrGly 472
 QY 1046 AAACCTCAAGGAGGAGAACTTTTGTGTCCTCCAGAGAGAAAGTGAAGCTGGAGACCCACATA 1105
 Db 473 LysLeuLysGluGluAsnPhePheGlyProLysGluGluValLysLeuGluThrHisIle 492
 QY 1106 CGTGTGCCAGCATCAGCAGCTGCGCGGTCTATTGGCAAGGTGAAAAACGGTGAACGAG 1165
 Db 493 ArgValProAlaSerAlaAlaGlyArgValIleGlyLysGlyGlyLysThrValAsnGlu 512
 QY 1166 TTGCAGAAATTTGACCGCAGCTGAGTGTAGTACCAAGACACAGACCCCTGATGAGAAC 1225
 Db 513 LeuGlnAsnLeuThrAlaAlaGluValValProArgAspGlnThrProAspGluAsn 532
 QY 1226 GACAGGTCATCTCGTGAATAATCATCGACATTTCTATCCAGTCAAGTGGCTCAACGGAAG 1285
 Db 533 AspGlnValIleValLysIleIleGlyHisPheTyrAlaSerGlnMetAlaGlnArgLys 552
 QY 1286 ATCCGAGACATCTCGGCCAGGTTAAAGCAGCAGCATCAGAAAGGACAGAGTAACAGGCC 1345
 Db 553 IleArgAspIleLeuAlaGlnValLysGlnGlnHisGlnLysGlyGlnSerAsnGlnAla 572
 QY 1346 CAGCACGAGGAGAAAG 1360
 Db 573 GlnAlaArgArgLys 577

RESULT 3
 O88477 PRELIMINARY; PRT; 577 AA.
 AC O88477;
 DT 01-NOV-1998 (TRENBLrel. 08, Created)
 DT 01-NOV-1998 (TRENBLrel. 08, Last sequence update)
 DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
 DE Coding region determinant binding protein.
 GN IGF2BP1 OR CRDBP.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]_TaxID=10090;
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92217743; PubMed=1559612;
 RA Bernstein P.L., Herrick D.J., Prokipcak R.D., Ross J.;
 RT "Control of c-myc mRNA half-life in vitro by a protein capable of
 RT binding to a coding region stability determinant.";
 RL Genes Dev. 6:642-654(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94158886; PubMed=8114742;
 RA Herrick D.J., Ross J.;
 RT "The half-life of c-myc mRNA in growing and serum-stimulated cells:
 RT influence of the coding and 3' untranslated regions and role of
 RT ribosome translocation.";
 RL Mol. Cell. Biol. 14:2119-2128(1994).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94179348; PubMed=8132663;
 RA Prokipcak R.D., Herrick D.J., Ross J.;
 RT "Purification and properties of a protein that binds to the C-terminal
 RT coding region of human c-myc mRNA";
 RL J. Biol. Chem. 269:9261-9269(1994).
 RN [4]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97322234; PubMed=9178888;
 RA Leeds P., Kren B.T., Boylan J.M., Betz N.A., Steer C.J.,
 RA Gruppiso P.A., Ross J.;
 RT "Developmental regulation of CRD-BP, an RNA-binding protein that
 RT stabilizes c-myc mRNA in vitro.";
 RL Oncogene 14:1219-1286(1997).
 RN [5]
 RP SEQUENCE FROM N.A.

QY	332	AGGAAGGAGAAACGGCAGGTCGACGTGAATAAAGCCATCAGTGTGCATCTCCACCCTCAGGGC	391
Db	233	ArgLysGluAsnAlaGlyValAlaGluuysAlaIleSerValHisSerThrProGluGly	252
QY	332	TGCTCTCCGTTGPAAGATGATCTTGAGATTATGCATAAAGAGCGTAAAGACACCAAA	451
Db	253	CysSerSerAlaCysLysMetCileLeuGluIleMetHisLysGluAlaLysAspThrLys	272
QY	452	ACGGCTGACGAGGTTCCCTCGAAGATCCTGCGCCCATATACTTTGTAGGGCGTCTCAAT	511
Db	273	ThrAlaAspGluValProLeuLysIleLeuAlaHisAsnAsnProValGlyArgLeuIle	292
QY	512	GGCAAGGAAGACCGAACTCTGAAGAAGGTAGACGAAGATACCGAGACAAAAATCACCATC	571
Db	293	GlyLysGluGlyArgAsnLeuLysLysValGluGlnAspThrGluThrLysIleThrIle	312
QY	572	TCCTCGTTGCAAGACCTTACCCTTTACAAACCTCGAGGAGCCATCAGTGAAGGGGGCC	631
Db	313	SerSerLeuGlnAspLeuThrLeuLysAsnProGluArgThrIleThrValLysGlyAla	332
QY	632	ATCAGAGAATTGTTCCAGGGCCGACGAGGAATAATGAAGAAGTTGGGAGGCCCTATGAG	691
Db	333	IleGluAsnCysCysArgAlaGluGlnGluIleMetLysLysValArgGluAlaArgGlu	352
QY	692	AATCATGTGGCTGCCATGAGC-----TCATCACTGATCCCTGGCCCTGAACCTGGCTGCT	745
Db	353	AsnAspValAlaAlaMetSerLeuGlnSerHisLeuIleProGlyLeuAsnLeuAlaAla	372
QY	746	GTAGGTCTTTTCCAGGTTTCATCAGCGCGAGTCCCGCGCTCCACGACGAGCTTACTGGG	805
Db	373	ValGlyLeuPheProAlaSerSerSerAlaValProProProProSerSerValThrGly	392
QY	806	GCTGCTCCCTATAGTCTCTTTATGCAGGCTCCCGAGCAGGAGATGGTCGAGTCTTTATC	865
Db	393	AlaAlaProTyrSerSerPheMetGlnAlaProGluGlnGluMetValGlnValPheIle	412
QY	866	CCGCCCAGCGAGTGGCGGCATCATCGGCAAGAGCGGACGACACATCAACAGCTCTCC	925
Db	413	ProAlaGlnAlaValGlyAlaIleIleGlyLysLysGlyGlnHisIleLysGlnLeuSer	432
QY	926	CGGTTTCCACGAGCGCTCCATCAAGATTGACACCCGAAACACCTGACTCCAAGTTCGT	985
Db	433	ArgPheAlaSerAlaSerIleLysIleAlaProProGluThrProAspSerLysValArg	452
QY	986	ATGGTTATCATCTGAGACCGCAGAGGCGCCAAATCAAGGCTCAGGAAAGAAATCTATGGC	1045
Db	453	MetValValIleThrGlyProProGluAlaGlnPheLysAlaGlnGlyArgIleTyrGly	472
QY	1046	AAACTCAAGCAGCAGAACCTCTTTGGTCCCAAGAGGAAGTGAAGTCGAGACCCACATA	1105
Db	473	LysLeuLysGluGlnAsnPhePheGlyProLysGluGluValLysLeuGluThrHisIle	492
QY	1106	CGTGTGCCAGCATCAGACGTGCGCGGTTCATTGGCAAAAGGTGGAATAACCGTGAAACGAG	1165
Db	493	ArgValProAlaSerAlaAlaGlyArgValIleGlyLysGlyLysThrValAsnGlu	512
QY	1166	TTCCAGAAATTGACGCAGCTGAGGTGGTAGTACCAGAGACACGACCCCTGATGAGAAC	1225
Db	513	LeuGlnAsnLeuThrAlaAlaGluValValProArgAspGlnThrProAspGluAsn	532
QY	1226	GACCAGGTCACTCGTGAANAATCATCGACATTTCTATGCCAGTCAGATGGCTCAACCGAAG	1285
Db	533	AspGlnValIleValLysIleIleGlyHisPheTyrAlaSerGlnMetAlaGlnArgLys	552
QY	1286	ATCCGAGACATCTGCGCCAGGTTAAGCAGCAGCATCAGAAGGACAGATACACGAGCC	1345
Db	553	IleArgAspIleLeuAlaGlnValLysGlnGlnHisGlnLysGlyGlnSerAsnLeuAla	572
QY	1346	CAGGCGCAGGAGAG 1360	
Db	573	GlnAlaArgArgLys 577	


```

RESULT 4
Q8BRH1 PRELIMINARY; PRT; 577 AA.
AC Q8BRH1;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Insulin-like growth factor 2.
GN IGF2BP1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J;
RX MEDLINE=22354683; PubMed=12466851;
RA The PANTOM Consortium.
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RL Nature 420:563-573 (2002).";
RL EMBL; AK044850; BAC32119.1; -.
DR MGD; MGI:1890357; Igf2bp1.
DR GO; GO:0003676; P:nucleic acid binding; IEA.
DR InterPro; IPR004087; KH_dom.
DR InterPro; IPR004088; KH_type_1.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00013; KH; 4.
DR Pfam; PF00076; rrm; 2.
DR SMART; SMC0360; RRM; 2.
DR SMART; SMC0322; KH; 4.
DR PROSITE; PS00084; KH_TYPE_1; 4.
DR PROSITE; PSS0102; RRM; 2.
SQ SEQUENCE 577 AA; 63378 MW; D75897450841532E CRC64;

Alignment Scores:
Pred. No.: 1,26e-158 Length: 577
Score: 2201.00 Matches: 437
Percent Similarity: 98.43% Conservative: 1
Best Local Similarity: 98.20% Mismatches: 5
Query Match: 70.77% Indels: 2
DB: 11 Gaps: 1

US-09-270-437D-5 (1-1708) x Q8BRH1 (1-577)
QY 32 CGGGGAGCCATCATGAAGCTGAATGGCCACAGTCAGTGGAGAACCATGCCCTGAGGTCCTCC 91
DB 133 ArgGlnAlaIleMetLysLeuAsnGlyHisGlnLeuGluAsnHisAlaLeuLysValSer 152
QY 92 TACATCCCGATGACAGATAGACACAGGACCTGAGAAATGGCGCCGAGGGGCTTTGGC 151
DB 153 TyrIleProaspGluGlnIleThrGlnGlyProGluasnGlyArgGlyGlyPheGly 172
QY 152 TCTCGGGTCAGCCCGCCAGGCTCACCTGTGGCAGCGGGGCCCGCCAGCAAGCAGCAG 211
DB 173 SerArgGlyGlnProArgGlnGlySerProValAlaAlaGlyAlaProAlaLysGlnGln 192
QY 212 CAAGTGACATCCCGCTCGCTCGCTCGTCCGCCACCCAGTATGTGGTGCCCATTTATGGC 271
DB 193 ProValAspIleProLeuArgLeuLeuValProThrGlnIleValGlyAlaIleIleGly 212
QY 272 AAGGAGGGGGCCACCATCCGCAACATCACAAAACAGACCCAGTCCCAAGATAGACGTGCAT 331
DB 213 LysGluGlyAlaThrIleArgAsnIleThrLysGlnThrGlnSerLysIleAspValHis 232
QY 332 AGGAAGAGAACCGAGGTGCGAGCTGAAAGCCATCAGTGTGCACTCCACCCCTGAGGCC 391
DB 233 ArgLysGluAsnAlaGlyAlaAlaGluLysAlaIleSerValHisSerThrProGluGly 252
QY 392 TGCTCTCCGCTTGTAGATCATCTTGAGATTATGCATAAAGAGGCTAAGGACACCAAAA 451
DB 253 CysSerSerAlaCysLysLysMetIleLeuGluIleMethisLysGluAlaLysAspThrLys 272

```

```

QY 452 ACGGCTGACGAGGTTCCCTGGAAGATCCTGGCCCATTAATTAACCTTTGAGGGCTCTCAT 511
DB 273 ThrAlaaspGluValProLeuLysIleLeuAlaHisAsnAsnPheValGlyArgLeuIle 292
QY 512 GGCAGGAAGACGAGAACCTGGAAGAAGTGAAGACATACCGAGACAAAATACCATC 571
DB 293 GlyLysGluGlyArgAsnLeuLysValGluGlnaspThrGluThrLysIleThrIle 312
QY 572 TCCTCGTTGCAAGACCTTACCTTTTACAACCTCGAGAGGACCATCATCTGTGAAGGGGGCC 631
DB 313 SerSerLeuGlnAspLeuThrLeuTyrAsnProGluArgThrIleThrValLysGlyAla 332
QY 632 ATCCGAGATTTGTCAGGGCCGACGAGCAATATATGAAGAAGTTCGGGAGGCTTATGAG 691
DB 333 IleGluAsnCysCysArgAlaGluGlnIleMetLysLysValAlaArgLysGlu 352
QY 692 AATGATGTGGCTGCATGAGC-----TCTCACCTGATCCCTGGCCTCAACCTGGCTGCT 745
DB 353 AsnAspValAlaAlaMetSerLeuGlnSerHisLeuIleProGlyLeuAsnLeuAlaAla 372
QY 746 GTAGTCTTTTCCCGAGCTTCATCCAGCGCAGTCCCGCGGCTCCCGAGGAGGTTACTGGG 805
DB 373 ValGlyLeuPheProAlaSerSerSerAlaValProProProSerSerValThrGly 392
QY 806 GGTGCTCCCTATAGCTCCCTTTATGAGGCTCCCGAGGAGGAGATGCTGAGGTGTTTATC 865
DB 393 AlaAlaProTyrSerSerPheMetGlnAlaProGluGlnGlyMetValGlnValPheIle 412
QY 866 CCGCCCGAGGAGTGGGGCCCATCATCCGCAAGAGGGGCGAGCATCAAAACAGCTCTCC 925
DB 413 ProAlaGlnAlaValGlyAlaIleIleGlyLysLysGlyGlnHisIleLysGlnLeuSer 432
QY 926 CGGTTTGGCAGGCTCCATCAAGATTGCACACCCGAAACACCTGACTCCAAAGTTCTGT 985
DB 433 ArgPheAlaSerAlaSerIleAlaProGluThrProaspSerLysValArg 452
QY 986 ATGGTTATCATCATCTGAGCCCGCCAGGAGCCCAATTCAGGCTCAGGGAGAAATCTATGC 1045
DB 453 MetValValIleThrGlyProGluAlaGlnPheLysAlaGlnGlyArgIleTyrGly 472
QY 1046 AAACATCAAGGAGGAGAACTTTCTTGGTCCCAAGGAGGAGTGAAGCTGAGAGCCACATA 1105
DB 473 LysLeuLysGluGluAsnPhePheGlyProLysGluGluValLysLeuGluThrHisIle 492
QY 1106 CQTGTGCAGCATCAGCAGCTGGCCGGCTCATTTGGCAAGGTGGAAAAACCGTGAACGAG 1165
DB 493 ArgValProAlaSerAlaAlaGlyArgValIleGlyLysGlyLysThrValAsnGlu 512
QY 1166 TTCCAGAATTTGACGGCAGCTGAGTGTGTAGTACCAAGAGACGACCCCTGATGAGAAC 1225
DB 513 LeuGlnAsnLeuThrAlaAlaGluValValProArgAspGlnThrProaspGluAsn 532
QY 1226 GACCAAGTTCATCGTGAATCATCGACATTTCTATGCGAGTCAGATGGCTCAACGGGAG 1285
DB 533 AspGlnValIleValLysIleIleGlyHisPheTyrAlaSerGlnMetAlaGlnArgLys 552
QY 1286 ATCCGAGACATCCTGGCCCGAGTTAAGCAGCAGCATCAAGGGGACAGAGTAAACGGGCC 1345
DB 553 IleArgAspIleLeuAlaGlnValLysGlnGlnHisGlnLysGlyGlnSerAsnLeuAla 572
QY 1346 CAGGACCGAGGAAG 1360
DB 573 GlnAlaArgArgLys 577

RESULT 5
Q80US9 PRELIMINARY; PRT; 577 AA.
AC Q80US9;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Igf2bp1 protein.

```

OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=129/Sv x 129/Sv-CP; TISSUE=Embryonic stem cells;
 RX MEDLINE=2238257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Stacenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,
 RA Bobak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 RA Krzywinski M.I., Skalska U., Smalios D.E., Schnerch A., Schein J.E.,
 RA Jones S.J., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RP [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=129/Sv x 129/Sv-CP; TISSUE=Embryonic stem cells;
 RA Strausberg R.;
 RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC051679; AHS1679.1; -;
 DR GO; GO:0003676; F:nucleic acid binding; IEA.
 DR InterPro; IPR004087; KH dom.
 DR InterPro; IPR004088; KH type 1.
 DR InterPro; IPR005054; RNA_rec_mot.
 DR Pfam; PF00013; KH; 4;
 DR Pfam; PF00076; Rrm; 2;
 DR SMART; SM00322; KH; 4;
 DR SMART; SM00360; RRM; 2;
 DR PROSITE; PSS0084; KH type 1; 4.
 DR PROSITE; PSS0102; RRM; 2;
 SQ SEQUENCE 577 AA; 63378 MW; D439F544257DA3CE CRC64;
 Alignment Scores:
 Pred. No.: 1,26e-158 Length: 577
 Score: 2201.00 Matches: 437
 Percent Similarity: 98.43% Conservative: 1
 Best Local Similarity: 98.20% Mismatches: 5
 Query Match: 70.77% Indels: 2
 DB: 11 Gaps: 1
 US-09-270-437D-5 (1-1708) x Q80U59 (1-577)
 QY 32 CGGGAGGCCATCATGAAGCTGAATGCCACCACAGTTGGAGAACCATGCCCTGAAGTCTCC 91
 Db 133 ArgGlnAlaIleMetLysLeuAsnGlyHisGlnLeuGluAsnHisAlaLeuLysValSer 152
 QY 92 TACATCCCGATGACAGATAGACAGGACCTGAGATGGCGCGCGGGGCTTTGGC 151
 Db 153 TyrIleProAspGluGlnIleThrGlnGlyProGluAsnGlyArgArgGlyGlyPheGly 172
 QY 152 TCTCGGGGTGAGCCCGCCAGGCTCACCTGTGGCAGCGGGGGCCCGCCAGCAGCAG 211
 Db 173 SerArgGlyGlnProArgGlnGlySerProValAlaAlaGlyAlaProAlaLysGlnGln 192
 QY 212 CAAGTGGACATCCCTTCGCTCTCTGGTGGCCACCCAGTAGTGGTGGCCATTATGCC 271
 Db 193 ProValAspIleProLeuArgLeuLeuValProThrGlnTyrValGlyAlaIleIleGly 212
 QY 272 AAGGAGGGGGCCACCATCCGCACATCACAAACAGACACCAGTCCCAAGTAGACGTGCAT 331

Db 213 LysGluGlyAlaThrIleArgAsnIleThrLysGlnThrGlnSerLysIleAspValHis 232
 QY 332 AGGAAGGAAACGCGAGTGCAGTGAAGAACCATCAGTGTGCACCTCCACCCCTGAGGCG 391
 Db 233 ArgLysGluAsnAlaGlyAlaAlaGluLysAlaIleSerValHisSerThrProGluGly 252
 QY 392 TCGCTCTCCGCTGTGAACATCATCTTGAGATTATGCATGAAGAGGCTAAGCAGCACCAA 451
 Db 253 CysSerSerAlaCysLysMetIleLeuGluIleMetHisLysGluAlaLysAspThrLys 272
 QY 452 ACGGTGACGAGGTTCCTCCAGATCCTGCGCCCATTAATCACTTTGTAGGCGCTCATTC 511
 Db 273 ThrAlaAspGlyValProLeuLysIleLeuAlaHisAsnAsnPheValGlyArgLeuIle 292
 QY 512 GGCAGGAGGACGCGACCTCAAGAGGTAGACAGATACCGAGACAGACAAATCACCATC 571
 Db 293 GlyLysGluGlyArgAsnLeuLysValGluGlnAspThrGluThrLysIleThrIle 312
 QY 572 TCCTCGTTTCAGACCTTACCCCTTTACACCCCTCAGAGGACCATCACTGTGAAGCGGCGC 631
 Db 313 SerSerLeuGlnAspLeuThrLeuTyrAsnProGluLarghThrIleThrValLysGlyAla 332
 QY 632 ATCGAGATTGTTGACGCGCGCGACAGCAATATATGAAGAAAGTTCGGGAGGCGCTATGAG 691
 Db 333 IleGluAsnCysCysArgAlaGluGlnGluIleMetLysLysValArgGluAlaTyrGlu 352
 QY 692 AATGATGTGGCTGCATGAGC-----TCTCACTGATCCTGCGCTGAACCTGGCTGCT 745
 Db 353 AsnAspValAlaAlaMetSerLeuGlnSerHisLeuIleProGlyLeuAsnLeuAlaIle 372
 QY 746 GTAGTCTTTTCCAGTTCATCCAGCGAGTCCCGCGCTCCAGCAGCGTACTTGGG 805
 Db 373 ValGlyLeuPheProAlaSerSerSerAlaValProProProProSerSerValThrGly 392
 QY 806 GCTCTCTCTATAGTCTCTTTATGCGAGTCCCGAGCAGGAGATGTTGCGAGGTGTTTATC 855
 Db 393 AlaAlaProTyrSerSerPheMetGlnAlaProGluGlnGluMetValGlnValPheIle 412
 QY 866 CCGCCCGAGCAGTGGCGCCCATCATCGCAAGAGGCGGAGCAGCACATCAACAGCTCTCC 925
 Db 413 ProAlaGlnAlaValGlyAlaIleIleGlyLysLysGlyGlnHisLysGlnLeuSer 432
 QY 926 CGGTTTGGCAGCGCTCCATCAAGATTGACACCCCGAACACACCTGACTCCAAAGTTCGT 985
 Db 433 ArgPheAlaSerAlaSerIleLysIleAlaProProGluThrProAspSerLysValArg 452
 QY 986 ATGTTTATCATCTGAGCCCGCAGAGGCCCAATTCAAGGCTCAGGAAAGAAATCTATGGC 1045
 Db 453 MetValValIleThrGlyProProGluAlaGlnPheLysAlaGlnGlyArgIleTyrGly 472
 QY 1046 AAATCAAGAGGAGAACTCTTTTGTGTCACAGAGAGAGTGAAGCTGGAGAGACCCACATA 1105
 Db 473 LysLeuLysGluGluAsnPhePheGlyProLysGluGluValLysLeuGluThrHisIle 492
 QY 1106 CGTGTGCGACATCAGCAGCTGCGCGCTCATTTGGCAAGGTGGAAGGTGGAAGGTGGAAG 1165
 Db 493 ArgValProAlaSerAlaAlaGlyArgValIleGlyLysGlyGlyLysThrValAsnGlu 512
 QY 1166 TTGCAGAAATTGACGCGAGCTGAGGTGGTAGTACCAGAGAGACAGACCCCTGATGAGAAC 1225
 Db 513 LeuGlnAsnLeuThrAlaAlaGluValValProArgAspGlnThrProAspGluAsn 532
 QY 1226 GACCAGTCTATCTGGAATCATCGGACATTTCTATGTCAGTGCAGATGCTGCTCAACGAG 1285
 Db 533 AspGlnValIleValLysIleIleGlyHisPheTyrAlaSerGlnMetAlaGlnArgLys 552
 QY 1286 ATCCGAGACATCTGCGCCAGGTTAAGCAGCAGCATCAGAGGAGCAGAGTAGTAACCAAGGC 1345
 Db 553 IleArgAspIleLeuAlaGlnValLysGlnGlnHisGlnLysGlyGlnSerAsnLeuAla 572
 QY 1346 CAGGCACGGAGGAG 1360

Db 573 GlnAlaArgArgLys 577

RESULT 6
O42254 PRELIMINARY; PRT; 576 AA.

AC O42254;
DT 01-JAN-1998 (T-EMBLrel. 05, Created)
DT 01-JAN-1998 (T-EMBLrel. 05, Last sequence update)
DE 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Zipcode-binding protein.
GN ZBP1.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinoptera; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus
OC NCB1_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9722007; PubMed=9121465;
RA Ross A.F., Olevnikov Y.S., Kislauskis E.H., Taneja K.L., Singer R.H.;
RT "Characterization of a beta-actin mRNA zipcode-binding protein.";
RL Mol. Cell. Biol. 17:2158-2165(1997).
CC -1- SIMILARITY: CONTAINS 4 KH DOMAINS.
DR EMBL; AF026527; AAB82295.1; -;
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR004087; KH dom.
DR InterPro; IPR004088; KH type 1.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00013; KH; 4.
DR Pfam; PF00076; rrm; 2.
DR SMART; SM00322; KH; 4.
DR SMART; SM00360; RRM; 2.
DR PROSITE; PS00084; KH TYPE 1; 4.
DR PROSITE; PS0102; RRM; 2.
DR PROSITE; PS00030; RRM_RNP 1; FALSE NEG
SQ SEQUENCE 576 AA; 63271 MW; 01A2F2D1D81C8911 CRC64;

Alignment Scores:
Pred. No.: 3,33e-152 Length: 576
Score: 2116.50 Matches: 419
Percent Similarity: 96.18% Conservative: 9
Best Local Similarity: 94.16% Mismatches: 14
Query Match: 68.05% Indels: 3
DB: 13 Gaps: 2

US-09-270-437D-5 (1-1708) x O42254 (1-576)

QY 32 CGGGAGCCATCATGAAGCTGAATGGCCACACAGTTGGAGAACCATGCCCTGAAGGTCTCC 91
Db 133 ArgGlnAlaLeuMetLysLeuAsnGlyHisGlnLeuGluAsnHisValLeuLysValSer 152
QY 92 TACATCCCGATGACGATAGCACAGGACCTGAGAAATCGGCGCCGAGGGGCTTTGGC 151
Db 153 TyrlleProaspGluGlnSerValGlnGlyProGluAsnGlyArgGlyGlyPheGly 172
QY 152 TCTCGGGGTAGCCCCCGCAGGGCTACCTGTGCGACGGGGGGCCCCAGCCAGCAGCAG 211
Db 173 AlaArgGlyAlaProArgGlnGlySerProValThrAlaGlyAlaProValLysGlnGln 192
QY 212 CAATGGACATCCCTTCGGCTCCTGGTGCACCCAGTATGTGGTGGCCATATTGGC 271
Db 193 ProValAspIleProLeuArgGluLeuValProThrGlnTyrValGlyAlaIleGly 212
QY 272 AAGAGGGGGCCACATCCGCAACATCACAAACAGACCCAGTCCCAAGATAGAGTGCAT 331
Db 213 LysGluGlyAlaThrIleArgAsnIleThrLysGlnThrGlnSerLysIleAspValHis 232
QY 332 AGGAGGAGGAGCAGGAGTGCAGTGAAGAAAGCCATCATGTGCTCCACCCCTGAGGGC 391
Db 233 ArgLysGluAsnAlaGlyAlaAlaGluLysAlaIleSerIleHisSerThrProGluGly 252
QY 392 TGCTCCTCCGCTTGAAGATCATCTTGAGATTATGCATAAAGAGGTAAAGGACACCAA 451

Db 253 CysSerAlaAlaCysLysMetIleLeuGluIleMetGlnLysGluAlaLysAspThrLys 272
QY 452 ACGGCTGACGAGTTCCTCCCTGAAGATCCTGGCCCATTAATACTTTGTAGGCGCTCATTT 511
Db 273 ThrAlaAspGluValProLeuLysIleLeuAlaHisAsnAsnPheValGlyArgLeuIle 292
QY 512 GCGAAGGAGGACGGAACCTGAAGAAGTGAAGCAAGATACCGAGACACAAAATCACCATC 571
Db 293 GlyLysGluGlyA-gaenLeuLysValGluGlnAspThrGluThrLysIleThrIle 312
QY 572 TCCTCGTTGCAAGACCTTACCTTTACACCTGAGAGGACCATCAGTGTGAAGGGGGCC 631
Db 313 SerSerLeuGlnAspLeuThrLeuTyrAsnProGluArgThrIleThrValLysGlySer 332
QY 632 ATCCGAGATTCTTCAGCGCGCAGCAGGAAATATGAAGAAAGTTCCGGAGGCGCTATGAG 691
Db 333 IleGluAsnCysCysLysAlaGluGlnIleMetLysValArgGluAlaTyrGlu 352
QY 692 AATGATGTGGTGCATGAGC-----TCTCACCTGATCCTCGCCTGAACCTGGCTGCT 745
Db 353 AsnAspValAlaAlaMetSerLeuGlnSerHisLeuIleProGlyLeuAsnLeuAlaAla 372
QY 746 GTAGGTCTTTCCAGCTTCATCCAGCGCAGTCCCGCGCTCCAGCAGAGGTTACTGGG 805
Db 373 ValGlyLeuPheProAlaSerSerAsnAlaValProProProSerSerValSerGly 392
QY 806 GCTGCTCCTTATAGCTCCTTTATGAGGCTCCCGAGCAGGAGATGGTGACAGTGTATC 865
Db 393 AlaAlaProTyrSerSerPheMet---ProProGluGlnGluThrValHisValPheIle 411
QY 866 CCGGCGCAGGAGTGGCGCCATCATCGGCAAGAGGGGCGACATCAACAGCTCTCC 925
Db 412 ProAlaGlnAlaValGlyAlaIleGlyLysLysGlyGlnHisIleLysGlnLeuSer 431
QY 926 CGGTTGCGAGCGCTCCATCAAGATTCCACACCCGAAACACCTGACTCCAAAGTTCGT 985
Db 432 ArgPheAlaSerAlaSerIleLysIleAlaProProGluThrProAspSerLysValArg 451
QY 986 ATGTTATCATCATGACCGGCGAGGCGCCAAATTCAGGCTCAGGGAAGAAATCTATGCG 1045
Db 452 MetValValIleThrGlyProProGluAlaGlnPheLysAlaGlnGlyArgIleTyrGly 471
QY 1046 ABACTCAAGGAGGAGAACTTCTTGTGCTCCCAAGGAGGAGTGAAGCTGGAGACCCACATA 1105
Db 472 LysLeuLysGluGluAsnPhePheGlyProLysGluGluValLysLeuGluThrIle 491
QY 1106 CGTGTGCCAGCATCAGCAGCTGCGCGGTCTATTGTCAGAGTGAAGAAACCGTGAACGAG 1165
Db 492 ArgValProAlaSerAlaAlaGlyArgValIleGlyLysGlyLysThrValAsnGlu 511
QY 1166 TTGCGAATTTGACGCGAGCTGAGTGTGCTAGTACCAAGAGACCCAGCCCTGATGAGAAC 1225
Db 512 LeuGlnAsnLeuThrAlaAlaGluValValProArgAspGlnThrProAspGluAsn 531
QY 1226 GACCAAGTTCATCGTAAAAATCATCGGACATTTCTATGCCAGTTCAGATGGCTCAACGGAAG 1285
Db 532 GluGlnValIleValLysIleIleGlyHisPheTyrAlaSerGlnMetAlaGlnArgLys 551
QY 1286 ATCCGAGACATCCTGCCCGAGGTTAAGCAGCAGCTCAGAGGAGCAGAGTACACGAGCC 1345
Db 552 IleArgAspIleLeuAlaGlnValLysGlnGlnHisGlnLysGlyGlnSerGlyGlnLeu 571
QY 1346 CAGGCGAGGAGGAG 1360
Db 572 GlnAlaArgArgLys 576

RESULT 7
O73932 PRELIMINARY; PRT; 594 AA.
ID O73932
AC O73932; (T-EMBLrel. 07, Created)
DT 01-AUG-1998 (T-EMBLrel. 07, Last sequence update)
DT 01-AUG-1998 (T-EMBLrel. 07, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)

DE Vg1 RNA binding protein variant D.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipiloidea; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RA Havin L., Git A., Elisha Z., Oberman F., Yaniv K.,
RA Pressman Schwartz S., Standart N.M., Yisraeli J.K.;
RL Genes Dev. 0:0-0(1998).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=98228351; PubMed=9560341;
RA Deshler J.O., Hight M.I., Abramson T., Schnapp B.J.;
RT "A highly conserved RNA-binding protein for cytoplasmic mRNA
RT localization in vertebrates.";
RL Curr. Biol. 8:489-496(1998).
CC - - SIMILARITY: CONTAINS 4 KH DOMAINS.
DR EMBL; AF064634; AAC18598.1; -.
DR EMBL; AF055923; AAC41285.1; -.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR004087; KH dom.
DR InterPro; IPR004086; KH_type_1.
DR InterPro; IPR005054; RNA_rec_mot.
DR Pfam; PF00013; KH; 4.
DR Pfam; PF00076; Rrm; 2.
DR SMART; SM00322; KH; 4.
DR SMART; SM00360; RRM; 2.
DR PROSITE; PS50084; KH TYPE_1; 4.
DR PROSITE; PS50102; RRM; 2.
SQ SEQUENCE 594 AA; 65643 MW; 54CEA7BF0856DD6 CRC64;
Alignment Scores:
Pred. No.: 8,43e-122 Length: 594
Score: 1716.50 Matches: 343
Percent Similarity: 84.52% Conservative: 50
Best Local Similarity: 73.76% Mismatches: 47
Query Match: 55.19% Indels: 25
DB: 13 Gaps: 7
US-09-270-437D-5 (1-1708) x 073932 (1-594)
QY 32 CGGGGACCATCATGAGTGAATGATGCCACCATGTCGAGAACATGCCCTGAAGTCTCC 91
DB 133 ArgGlnGlyLeuGluLysLeuAsnGlyTyrGlnLeuGluAsnTyrSerLeuLysValThr 152
QY 92 TACATCCCCGATGAG-----CAGATGACAGGAGCCT 124
DB 153 TyrIleProAspGluMetAlaThrProGlnAlaProSerGlnGlnLeuGlnGlnPro 172
QY 125 GAG-----AATGGGCCCGAGGGGGCTTGGCTCTCGGGGTGAGCCCGCCAG 172
DB 173 GlnGlnGlnHisProGlnGlyArgArg---GlyPheGlyGlnArgGlyProAlaArgGln 191
QY 173 GGCTCACCTGTGCGAGCGGGGGCCAGCAAGCAGCAGCAGTGCATCCCTCGG 232
DB 192 GlySerProGlyAlaAlaAlaArgPro-----LysProGlnThrGluValProLeuArg 209
QY 233 CTCTGTGTCGCCACCATGATGTTGGTGCATATTGCAAGAGGGGGCCACCATCCGC 292
DB 210 MetLeuValProThrGlnPheValGlyAlaIleGlyLysGluGlyAlaThrIleArg 229
QY 293 AACATCACAACAGACCCAGTCCAGATAGACGTGCATAGGAGGAGAACCGAGTGCA 352
DB 230 AsnIleThrLysGlnThrGlnSerLysIleAspIleHisArgGlyGluAsnAlaGlyAla 249
QY 353 GCTGAATAAGCCATGATGTGCACTCCACCTGAGGGTGTCTCTCGCTTGTGAAGATG 412
DB 250 AlaGluLysProIleThrIleHisSerThrProGluGlyCysSerAlaAlaCysLysIle 269
QY 413 ATCTTGAGATATGATGAAGAGGTAGGACACCAACCGCTGACGAGGTCCCTCG 472

DB 270 IleMetGluIleMetGlnLysGluAlaGlnAspThrLysPheThrGluGluIleProLeu 289
QY 473 AAGATCTCTGGCCCAATAAATCTTTAGGGCGTCTCATTTGGCAAGGAGCGGAACCTG 532
DB 290 LysIleLeuAlaHisAsnAsnPheValGlyArgLeuIleGlyLysGluGlyArgAsnLeu 309
QY 533 AGRAGGTAGAGAGATACCGAGACAAAATACCATCTCTCTCTCTCTCTCTCTCTCTCT 592
DB 310 LysLysIleGluGlnAspThrLysIleThrIleSerProLeuGlnAspLeuThr 329
QY 593 CTTTACAAACCTGAGAGACCATCATCTGTGAAGGGGGCCATCGAGAATTTGTTCAGGGCC 652
DB 330 LeuTyrAsnProGluArgThrIleThrValLysGlySerIleGluProCysAlaLysAla 349
QY 653 GAGCAGGAATAATGAGAAAGTTCGGAGGCCCTATGAGATGATGATGATGATGATGATG 712
DB 350 GluGluGluIleMetLysLysIleArgGluSerTyrGluAsnAspIleAlaAlaMetAsn 369
QY 713 -----TCTACCTGATCCCTGGCTGAACCTGGCTGCTGTAGTCTTTTCCAGCTTCA 766
DB 370 LeuGlnAlaHisLeuIleProGlyLeuAsnLeuAsnAlaLeuGlyLeuPheProSerSer 389
QY 767 TCCAGCGAGTCCCGCGCTCCCGAGCAGGCT-----ACTGGGGTGTCTCCCTAT 817
DB 390 SerSerGlyMetProProProSerValGlyValProSerProThrSerSerThrSerTyr 409
QY 818 AGCTCTTTATGTCAGGCTCCCGAGCAGGATGTCAGGCTTTATCCCGCCAGGCA 877
DB 410 ProProPheGlyGlnGlnProGluSerGluThrValHisLeuPheIleProAlaLeuAla 429
QY 878 GTGGCGCCATCATCGCAAGAGGGGCGACACATCAACAGCTCTCCCGTTTGGCAGC 937
DB 430 ValGlyAlaIleIleGlyLysGlnHisIleLysGlnLeuSerArgPheAlaGly 449
QY 938 GCCTCCATCAGATTGCACACCCGAAACACCTGACTCCAAAGTCTCGTATGTTATCATC 997
DB 450 AlaSerIleLysIleAlaProAlaGluGlyProAspAlaLysLeuArgMetValIle 469
QY 998 ACTGGACCGCCAGAGGCCAATTCAGGCTCAGGAGAAATCTATGGCAAACTCAAGGAG 1057
DB 470 ThrGlyProProGluAlaGlnPheLysAlaGlnArgIleTyrGlyLysLeuLysGlu 489
QY 1058 GAGAACTTCTTGTCTCCAGGAGGAAGTGAAGCTGGAGACCCACATACATGTGCGCAGCA 1117
DB 490 GluAsnPheGlyProLysGluValLysLeuGluThrHisIleLysValProSer 509
QY 1118 TCAGCAGTGGCGGGTCATTGGCAAGGTGGAAGAAACCGTGAACGAGTTGCAGAAATTG 1177
DB 510 TyrAlaAlaGlyArgValIleGlyLysGlyLysThrValAsnGluLeuGlnAsnLeu 529
QY 1178 AGCGACGTGAGTGTGTAGTACCAAGAGACAGACCCCTCATGAGAACGACAGGTCATC 1237
DB 530 ThrSerAlaGluValValProArgAspGlnThrProAspGluAsnAspGluValVal 549
QY 1238 GTGAAAATCATCGACATTTCTATGCGAGTCAGATGGCTCAACGAGAGATCCGAGACATC 1297
DB 550 ValLysIleThrGlyHisPheTyrAlaSerGlnLeuAlaGlnArgLysIleGlnGluIle 569
QY 1298 CTGCCCCAGGTTAAGCAGCAGCATCAGAGGGA-----CAGAGTAACACGAGGCC 1345
DB 570 LeuAlaGlnValArgArgGlnGlnGlnGlnGlnLysThrValGlnSerGlyGlnPro 589
QY 1346 CAGCAGCGAGGAG 1360
DB 590 GlnProArgArgLys 594
RESULT 8
ID O57526 PRELIMINARY; PRT; 593 AA.
AC O57526;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE KH domain-containing transcription factor B3.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
 OC Xenopodinae; Xenopus.
 OX NCBI_TaxID=8355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92249652; PubMed=1577195;
 RA Pfaff S.L., Taylor W.L.;
 RT "Characterization of a Xenopus oocyte factor that binds to a
 RT developmentally regulated cis-element in the TFIIIA gene.";
 RL Dev. Biol. 151:306-316(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Griffin D., Taylor W.L.;
 RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Havin L., Git A., Elisha Z., Oberman F., Yaniv K.,
 RA Pressman Schwartz S., Standart N.M., Yisraeli J.K.;
 RL Genes Dev. 0:0-0(1998).
 CC -1- SIMILARITY: CONTAINS 4 KH DOMAINS.
 DR EMBL; AF042353; AAB97457.1; -.
 DR EMBL; AF064633; AAC18597.1; -.
 DR GO; GO:0003676; F:nucleic acid binding; IEA.
 DR InterPro; IPR004087; KH dom.
 DR InterPro; IPR004088; KH_type_1.
 DR InterPro; IPR000504; RNA_rec_mot.
 DR Pfam; PF00013; KH; 4.
 DR Pfam; PF00076; rrm; 2.
 DR SMART; SM00322; KH; 4.
 DR SMART; SM00360; RRM; 2.
 DR PROSITE; PS50084; KH TYPE_1; 4.
 DR PROSITE; PS50102; RRM; 2.
 DR PROSITE; PS00030; RRM_RNP_1; FALSE_NEG.
 DR PROSITE; PS00030; RRM_RNP_1; FALSE_NEG.
 SQ SEQUENCE 593 AA; 65385 MW; 5A5AB4B4A1D55D7 CRC64;
 Alignment Scores:
 Pred. No.: 9,26-122 Length: 593
 Score: 1716.00 Matches: 345
 Percent Similarity: 84.27% Conservative: 46
 Best Local Similarity: 74.35% Mismatches: 49
 Query Match: 55.18% Indels: 24
 DB: 13 Gaps: 7
 US-09-270-437d-5 (1-1708) x 057526 (1-593)
 QY 32 CGGGGACCATCATGAGCTGAATGGCCACAGTGGAGAACCATGCGCTGAGGTCCTC 91
 Db 133 ArgGlnGlyLeuGluLysLeuAsnGlyTy-GlnLeuGluAsnTyrSerLeuLysValThr 152
 QY 92 TACATCCCGATGAG-----CAGATAGCACAGGGACCTGAG 127
 Db 153 TyrIleProAspGluMetAlaThrProGlnSerProSerGlnGlnLeuGlnProGln 172
 QY 128 -----AATGGGCGCGAGGGGCTTGGCTCTCGGGTACGCCCCCGAGGC 175
 Db 173 GlnGlnHisProGlnGlyArg- --GlyPheGlyGlnArgGlyProAlaArgGlnGly 191
 QY 176 TCACCTGTGGAGCGGGGGCCCCAGCAAGCAGCAGCAAGTGGACATCCCGCTCCGCTC 235
 Db 192 SerProGlyAlaAlaAlaArgProLysProGlnSerGlnVal-----ProLeuArgMet 209
 QY 236 CTGTGTGCCCCAGTATGGGTGGCCATTATTGGCAAGGAGGGGCCACCATCCGCAAC 295
 Db 210 LeuValProThrGlnPheValGlyAlaIleIleGlyLysGluGlyAlaThrIleArgAsn 229
 QY 296 ATCAAAACAGACCCAGTCCAGATAGACGTGCATAGGAGGAGGACGCGAGGTGCAGCT 355
 Db 230 IleThrLysGlnThrGlnSerLysIleAspIleHisArgLysGluAsnAlaGlyAlaAla 249
 QY 356 GAAAAAGCCATCAGTGTGCATCCACCCCTGAGGGCTGCTCCTCGCTTGTAAAGATGATC 415

RESULT 9
 000425

Db 250 GluLysProIleThrIleHisSerThrProGluGlyCysSerAlaAlaCysLysIleIle 269
 QY 416 TTGGAGATTATGCATAAAGAGCTTAAGACACCAAAACGGCTGACGAGGTTCCTCTGAAG 475
 Db 270 MetGluIleMetGlnLysGluAlaGlnAspThrLysPheThrGluGluIleProLeuLys 289
 QY 476 ATCTGGGCCCATATAACTTTGTAGGGCTCTCAATGGCAAGGAGGACGGAACCTGAAG 535
 Db 290 IleLeuAlaHisAsnAsnPheValGlyArgLeuIleGlyLysGluGlyArgAsnLeuLys 309
 QY 536 AAGGTAGACAGATACCGACACAAAAATCACCATCTCTCTCGTTCGCAAGACCTTACCTT 595
 Db 310 LysIleGluGlnAspThrAspThrLysIleThrIleSerProLeuGlnAspLeuThrLeu 329
 QY 596 TACACCCCTGAGAGACCATCACTGTGAGGGGCCCATCGAGAATTGTTGAGGGCCGAG 655
 Db 330 TyrAsnProGluArgThrIleThrValLysGlySerIleGluThrCysAlaLysAlaGlu 349
 QY 656 CAGGAATAATGAGAAAGTTCCGGAGCGCTTATGAGATGATGTGGCTGCCATGAGC--- 712
 Db 350 GluGluValMetLysLysIleArgGluSerTyrGluAsnAspIleAlaAlaMetAsnLeu 369
 QY 713 ---TCTCACTGATCCCTGGCTGACCTGAGCTGCTGTAGGCTCTTCCGAGCTTCATCC 769
 Db 370 GlnAlaHisLeuIleProGlyLeuAsnLeuAlaLeuGlyLeuPheProProSerSer 389
 QY 770 AGCGCAGTCCCGCGCTCC-----AGCAGCGTTACTGGGGCTGCTCCCTATAGC 820
 Db 390 SerGlyMetProProProSerAlaGlyValSerSerProThrThrSerAlaSerTyrPro 409
 QY 821 TCCTTTATGAGGCTCCCGAGCAGAGATGGTGCAGGTGTATTATCCCGCCGAGCAGTG 880
 Db 410 ProPheGlyGlnGlnProGluSerGluThrValHisLeuPheIleProAlaLeuAlaVal 429
 QY 881 GCGCCCATCTCGCAAGAGGGGCGAGCACATCAACAGCTCTCCCGGTTGCCAGCGCC 940
 Db 430 GlyAlaIleIleGlyLysGlnGlyHisIleLysGlnLeuSerArgPheAlaGlyAla 449
 QY 941 TCCATCAAGATTGCACACCCGAAACACCTGACTCCAAAGTTCGTATGTTATCATCACT 1000
 Db 450 SerIleLysIleAlaProAlaGluGlyProAspAlaLysLeuArgMetValIleIleThr 469
 QY 1001 GGACCGCAGAGCGCCCAATTCAAGCTCAGGAGAAATCTATGSCAAACTCAAGGAGG 1060
 Db 470 GlyProProGluAlaGlnPheLysAlaGlnGlyArgIleTyrGlyLysLeuLysGluGlu 489
 QY 1061 AACTTCTTTGGTCCCAAGGAGGAAGTGAAGCTGGAGACCCCATACGTGTGCCAGCATCA 1120
 Db 490 AsnPhePheGlyProLysGluGluValLysLeuGluAlaHisIleLysValProSerTyr 509
 QY 1121 CGAGCTGGCGGGCTCATTTGGCAAGGTGGAAGAAACGGTGAACGAGTTCAGAGATTGAGC 1180
 Db 510 AlaAlaGlyArgValIleGlyLysGlyGlyThrValAsnGluLeuGlnAsnLeuThr 529
 QY 1181 CGAGCTGAGGTGATAGTACCAAGACACAGACCCCTGATGAGAACACGACAGTCACTG 1240
 Db 530 SerAlaGluValValValProArgAspGlnThrProAspGluAsnAspGlnValVal 549
 QY 1241 AAAATCATCGACATTTCTATCCAGTCAGTGGTTCACCGGAAGATCCGAGACATCTG 1300
 Db 550 LysIleThrGlyHisPheTyrAlaSerGlnLeuAlaGlnArgLysIleGlnGluIleLeu 569
 QY 1301 GCCCAGGTTAACGACGACATCAGAAG-----GGACAGGTAAACAGGCCGAG 1348
 Db 570 AlaGlnValArgArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 589
 QY 1349 GCACGAGGAAG 1360
 Db 590 ProArgArgLys 593

DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE 10 days embryo cDNA, RIKEN full-length enriched library,
DE clone:2610036348, full insert sequence (Igf2 mRNA-binding protein 3)
DE (insulin-like growth factor 2, binding protein 3).
GN IGFBP3 OR 2610101N11K1K OR MIM33.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Embryo;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staebli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Monbaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez J., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.
RL Functional annotation of a full-length mouse cDNA collection.;
RL Nature 409:685-690(2001).
RN [2]
RP SEQUENCE FROM N.A.
RA Mori H., Sakakibara S., Imai T., Nakamura Y., Iijima T., Suzuki A.,
RA Yuasa Y., Takeda M., Okano H.;
RT "Expression of mouse Igf2 mRNA-binding protein 3 and its implications
RT for the developing central nervous system.";
RL J. Neurosci. Res. 0:0-0(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain, and Olfactory epithelium;
RX MEDLINE=22388257; PubMed=12477932;
RA Klausner R.D., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins P.S., Wagner L., Siemsen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Besak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villaion D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain, and Olfactory epithelium;
RA Strausberg R.;
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: CONTAINS 4 KH DOMAINS.
DR EMBL; AK011689; BAB27779.1; -.

DR EMBL; AB046173; BAB19755.1; -.
DR EMBL; BC045138; AAB45138.1; -.
DR EMBL; BC049082; AAB49082.1; -.
DR MGD; MGI:1890359; Igfbp3.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR004087; KH dom.
DR InterPro; IPR004088; KH_type_1.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00013; KH; 4.
DR Pfam; PF00076; rrm; 2.
DR SMART; SM00322; KH; 4.
DR SMART; SM00360; RRM; 2.
DR PROSITE; PS00084; KH_Type_1; 4.
DR PROSITE; PS0102; RRM; 2.
SQ SEQUENCE 579 AA; 63574 MW; CABD9A435B392B7 CRC64;
Alignment Scores:
Pred. No.: 4,45e-115 Length: 579
Score: 1628.00 Matches: 334
Percent Similarity: 83.92% Conservative: 47
Best Local Similarity: 73.57% Mismatches: 55
Query Match: 52.35% Indels: 18
DB: 11 Gaps: 10
US-09-270-437D-5 (1-1708) x Q9CPN8 (1-579)
QY 32 CGGGAGGACCATCATGAAGCTGAATGGCCACAGTTGGAGAACCATGCTCGAAGGTCTCC 91
Db 133 ArgGlnAlaLeuAspLysLeuAsnGlyPheGlnLeuGluAsnPheThrLeuLysValAla 152
QY 92 TACATCCCGCATGAGCAGATGACAG-----GGACCTGAG---AATGGGCGCGA 139
Db 153 TyrIleProAspGluThrAlaAlaGlnGlnAsnProSerProGlnLeuArgGlyArgArg 172
QY 140 GGGGGCTTTGGCTCTCGGGGTCAGCCCGCCAGGGCTCACCTGTGGCAGCGGGGCCCA 199
Db 173 GlyPro---GlyGlnArgLysSerArgGlnAlaSerPro-----GlySerVal 188
QY 200 GCACAGCAGCAGCAAGTGCATCCCTCGGTCTCGGTGCTCGGCGCCACCATGATGTGGT 259
Db 189 SerLysGlnLysProCysAspLeuProLeuArgLeuLeuValProThrGlnPheValGly 208
QY 260 GCATTTATGGCAGAGGGGGCCACCATCCGCAACATCAGCAACACACACACAGCTCCAG 319
Db 209 AlaIleIleGlyLysGluGlyAlaThrIleArgAsnIleThrLysGlnThrGlnSerLys 228
QY 320 ATGACGCTGCATAGGAGGAGAACCGAGTGCAGCTGAGTGAAGAACCATCAGTGTGCATCC 379
Db 229 IleAspValHisArgLysGluAsnThrGlyAlaAlaGluLysSerIleThrIleLeuSer 248
QY 380 ACCCTGAGGGTGTCTCTCCGCTTGTAAATGATCTTTGGAGATTATGCAATAAGAGGCT 439
Db 249 ThrProGluGlyThrSerAlaAlaCysLysSerIleLeuGluIleMetHisLysGluAla 268
QY 440 AAGGACACCAACCGCTGACGAGGTTCCCTGAGATCCCTGGCCCAATAAATCTTGA 499
Db 269 GlnAspIleLysPheThrGluGluIleProLeuLysIleLeuAlaHisAsnAsnPheVal 288
QY 500 GGGCGCTCTCATGGCAAGGAGCGAACCTGAGAGAGGTAGACCAAGATACCGAGACA 559
Db 289 GlyArgLeuIleGlyLysGluGlyArgAsnLeuLysLysIleGluGlnAspThrAspThr 308
QY 560 AAAATACCATCTCTCTCGTTCAGACACCTTACCTTTTCAACCTTTGAGAGGACCATCACT 619
Db 309 LysIleThrIleSerProLeuGlnGluLeuThrLeuTyrAsnProGluArgThrIleThr 328
QY 620 GTGAAGGGGGCCATCCAGAAATTTGTGACGGGCGGAGCAAGAAATAATAAGAAAGTTCGG 679
Db 329 ValLysGlySerValGluThrCysAlaLysAlaGluGluIleMetLysLysIleArg 348
QY 680 GAGGCGCTAGGAATGATGTGGTCCCATGAGC-----TCTCACTGATCTCGGCTG 733
Db 349 GluSerTyrGluAsnAspIleAlaSerMetAsnLeuGlnAlaHisLeuIleProGlyLeu 368

QY 734 AACCTGCTGCTAGTCTTTTCCAGCTTCATCCAGCGAGTCCGCGC----- 784
 Db AsnLeuAsnAlaLeuGlyLeuPheProProThrsSerGlyMetProProThrsSerGly 388
 QY 785 CCTCCAGCAGCGTTACTGGGGTCTCCCTATAGCTCCTTTATGAGCGCTCCCGAGCAG 844
 Db ProProSerThrsLeuThr-----ProProTy-ProGlnPheGluGlnSer---GluThr 405
 QY 845 GAGATGTGTGAGGNGTTTATCCCGCCAGCGAGTGGCGCCATCTCCGACAGAGGG 904
 Db GluThrValHisLeuPheLeuAlaLeuSerValGlyAlaIleGlyLeuGlnGly 425
 QY 905 CAGCACATCAAAAGCTCTCCCGTTTCCAGCGCTCCATCAAGATTGACACACCGAA 964
 Db GlnHisLeuGlnLeuSerArgPheAlaGlyAlaSerIleLeuAlaProAlaGlu 445
 QY 965 ACACCTGACTCCAAAGTTCGTATGTTTATCATCTGACCGCGCAGAGGCCAATTCAAG 1024
 Db AlaProAspAlaLeuValArgMetValIleIleThrGlyProGluAlaGlnPheLys 465
 QY 1025 GCTCAGGGAAGTCTATGTCACAACTCAAGGAGGAGAACTTCTTGTCTCCAGAGGAA 1084
 Db AlaGlnGlyArgIleTyrglyLysIleLysLeuGluAsnPheValSerProLysGluGlu 485
 QY 1085 GTGAAGCTGGAGACCCATACGTGTGCCAGCATCAGCGAGTGGCGGGTCAATGGCAA 1144
 Db ValLysLeuGluAlaHisIleArgValProSerPheAlaGlyArgValIleGlyLys 505
 QY 1145 GGTGAAAGACGGTGAACGAGTTGTCAGAAATTGACGCGAGTGGTGTAGTACCAAGA 1204
 Db GlyGlyLysThrValanGluLeuGlnSerLeuSerAlaGluValValProArg 525
 QY 1205 GACGAGCCCTGATGAGACGACCGAGTTCATCGTGAATATCGGACATTTCTATGCC 1264
 Db AspGlnThrProAspGluAsnArgGlnValValLysIleThrGlyHisPheTyra 545
 QY 1265 AGTCAGATGCTCAACGGAGATCCGAGACATCTGCGCCAGGTTAAGCAG--CAGCAT 1321
 Db CysGlnValAlaGlnArgLysIleGlnGluIleLeuThrGlnValLysGlnHisGlnGln 565
 QY 1322 CAGAAGGGA---CAGAGTAACCGCGCCAGCGAGGAGG 1360
 Db GlnLysAlaLeuGlnSerGlyProProGlnSerArgArgLys 579

RESULT 11
 Q8C2J9 PRELIMINARY; PRT; 579 AA.
 AC Q8C2J9
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Insulin-like growth factor 2.
 GN IGF2BP3.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NOD; TISSUE=Thymus;
 RX MEDLINE=22354683; PubMed=12466851;
 RA The FANTOM Consortium,
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs."
 RL Nature 420:563-573(2002).
 DR ENBL; AK08465; BAC40370.1; -.
 DR MGD; MGI:1890359; Igf2bp3.
 DR GO; GO:0003676; F:nicotinic acid binding; IEA.
 DR InterPro; IPR004087; KH dom.
 DR InterPro; IPR004088; KH type 1.
 DR InterPro; IPR000504; RNA_rec_mot.

DR Pfam; PF00013; KH; 4.
 DR Pfam; PF00076; xrm; 2.
 DR SMART; SM00322; KH; 4.
 DR SMART; SM00360; RRM; 2.
 DR PROSITE; PS50084; KH TYPE 1; 4.
 DR PROSITE; PS50102; RRM; 2.
 SQ SEQUENCE 579 AA; 63551 MW; 937E601A95D06B77 CRC64;

Alignment Scores:
 Pred. No.: 1,51e-114 Length: 579
 Score: 1621.00 Matches: 333
 Percent Similarity: 83.92% Conservative: 48
 Best Local Similarity: 73.35% Mismatches: 55
 Query Match: 52.12% Indels: 18
 DB: 11 Gaps: 10

US-09-270-437D-5 (1-1708) x Q8C2J9 (1-579)

QY 32 CGGGAGCCATCATGAAGCTGAATGGCCACAGTTGGAGAACCATGCCCTGAAAGTCTCC 91
 Db ArgGlnAlaLeuAspLysLeuAsnGlyPheGlnLeuGluAsnPheThrLeuLysValAla 152
 QY 92 TACATCCCGATGACGATAGCACAG-----GGACCTGAG---AATGGCGCGCA 139
 Db TyrIleProAspGluThrAlaAlaGlnGlnAsnProSerProGlnLeuArgGlyArg 172
 QY 140 GGGGGCTTTGGCTCTCGGGGTTCAGCCCGCCAGGGCTCACCTGTGGCAGCGGGGGCCCA 199
 Db GlyPro---GlyGlnArgGlySerSerArgGlnAlaSerPro-----GlySerVal 188
 QY 200 GCCAAGCAGCAGCAAGTGGACATCCCTTCGGCTCTGGTCCGCCACCCAGTATGCTGGT 259
 Db SerLysGlnLysProCysAspLeuProLeuArgLeuValProThrGlnPheValGly 208
 QY 260 GCCATTATTGGCAAGGAGGGGCCACCATCGCAACATCAAAACAGACCCAGTCCAAG 319
 Db AlaIleIleGlyLysGluGlyAlaThrIleArgAsnIleThrLysGlnThrGlnSerLys 228
 QY 320 ATAGAGTGCATAGAGAGAGACGCGAGTGCAGCTGAGCTGAGAAAGCCATCAGTGTGCTCC 379
 Db IleAspValHisArgLysGluAsnThrGlyAlaAlaGluLysSerIleThrIleLeuSer 248
 QY 380 ACCCTCAGGGCTCTCTCTCGCTTGAAGATGATCTTGGAGATTATGCAATAAAGAGGCT 439
 Db ThrProGluGlyThrSerAlaAlaCysLysSerIleLeuGluIleMetHisLysGluAla 268
 QY 440 AAGGACACCAACCGCTGACGAGTTCCTCGAAGATCTCGGCCCAATAATACTTTGTA 499
 Db GlnAspIleLysPheThrGluGluIleProLeuLysIleLeuAlaHisAsnAsnPheVal 288
 QY 500 GGGCGTCTCATTTGGCAAGGAGGAGCAACTGAGAGGTTAGAGCAAGATACCGAGACA 559
 Db GlyArgLeuIleGlyLysGluGlyArgAsnLeuLysIleGluGlnAspThrAspThr 308
 QY 560 AAAATCACCATCTCTCTGTTGCAAGACCTTACCTTTACACCTGAGAGGACCATCCT 619
 Db LysIleThrIleSerProLeuGlnGluLeuThrLeuTyraAsnProGluArgThrIleThr 328
 QY 620 GTGAGGGGGCCCATCGAAGATTGTCAGGGCCGAGCAGGAGAAATATGAAGAAGTCCG 679
 Db ValLysGlySerValGluThrCysAlaLysAlaGluGluIleMetLysIleArg 348
 QY 680 GAGCGCTATGAGATGATGTGGTGCCTCATGAGC-----TCTCACCTGTATCCCTGGCTG 733
 Db GluSerTyrgluAsnAspIleAlaSerMetAsnLeuGlnAlaAsnLeuIleProGlyLeu 368
 QY 734 AACCTGCTGCTAGTCTTTTCCAGCTTCATCCAGCGCAGTCCCGCGC----- 784
 Db AsnLeuAsnAlaLeuGlyLeuPheProProThrsSerGlyMetProProThrsSerGly 388
 QY 785 CCTCCAGCAGCGTTACTGGGGTCTCCCTATAGCTCCTTTATGAGCGCTCCCGAGCAG 844
 Db ProProSerThrsLeuThr-----ProProTy-ProGlnPheGluGlnSer---GluThr 405


```

Qy 632 ATCAGAGATTGTTGTCAGGGCCGAGCGAGGAATAATGAAGAAGTTCGGGAGGCCTATGAG 691
   :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 331 ValGluAlaCysAlaSerAlaGluIleGluIleMetLysLysLeuArgGluAlaPheGlu 350
   :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 692 AATGATGTCGTCCTGACAGCTCTCACTGATCCCTGGCTGAACTGGCTGCTGAGGT 751
   :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 351 AsnAspMetLeuAlaValAsnThrHis-----SerGly 361
   :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 752 CTTTTCACAGCTTCATCCAGCGAGTCCCGCCGCTCCAGCAGCGCTTACTGGGGTGTG 811
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 362 TyrPhe-----SerSerLeuTyrProHisHisGln-----PheGly 373
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 812 CCTATAGCTCCTTATGCAAGCTCCCGACGAGGATGTGTCAGGTGTTATCCCGGCC 871
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 374 ProPheProHisHisSerTyrProGluGluIleValAsnLeuPheIleProThr 393
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 872 CAGCGAGTGGCGCCATCATCGGCAAGAGGGGAGCAGCATCAAAAGCTCTCCCGGTTT 931
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 394 GluAlaValGlyAlaIleIleGlyLysLysGlyAlaHisIleLysGlnLeuAlaArgPhe 413
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 932 GCCAGCGCTCCATCAAGATTGACACCCGGAACACCTGACTCCAAAGTTCGTATGTT 991
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 414 AlaGlyAlaSerIleLysIleAlaProAlaGluGlyProAspValSerGluArgMetVal 433
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 992 ATCATCACTGCAGCGCCGAGGCGCAATTCAAGGCTCAGGGAAGAATCTATGCAGAACTC 1051
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 434 IleIleThrGlyProProGluAlaGlnPheLysAlaGlnGlyArgIlePheGlyLysLeu 453
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1052 AAGAGAGAACTTTCTTGTGTCCTCAAGAGGAGAGTGAAGCTGGAGACCCACATACGTGG 1111
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 454 LysGluGluAsnPhePheAsnProLysGluGluValLysLeuGluAlaHisIleArgVal 473
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1112 CCAGCATCAGCAGCTGCGCGGGTCATTGGCAAGGTGGAAGAAACGGTGAACGAGTTCGAG 1171
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 474 ProSerThrAlaGlyArgValIleGlyLysGlyLysThrValAsnGluLeuGln 493
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1172 AATTGACGCGAGCTGAGGTGGTAGTACCAAGAGACAGACCCCTGATGAGAACGACCGAG 1231
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 494 AsnLeuThrSerAlaGluValIleValProArgAspGlnThrProAspGluAsnGluGlu 513
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1232 GTCATCTGAAATCATCGACATCTTCTATCCAGTCAGATGCTCAACGGAGATCCGA 1291
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 514 ValIleValArgIleIleGlyHisPheAlaSerGlnThrAlaGlnArgLysIleArg 533
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1292 GACATCTGCGCCAGGTTAAGCAGCAGCATCAGAGGGACAGAGTAACACGAGCGCCAGGCA 1351
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 534 GluIleValGlnGlnValLysGlnGluGlnLysTyrProGlnGlyValAlaSerGln 553
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1352 CGGAGGAAG 1360
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 554 ArgSerLys 556
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 14
Q7TQF9
ID Q7TQF9 PRELIMINARY; PRT; 545 AA.
AC Q7TQF9;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
ON NCBI_TaxID=10090;
RX STRAIN=C57BL/6; TISSUE=Brain;
RX MEDLINE=2238257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,

```

```

RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Villalón D.K., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Richardson D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickinson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RA Strausberg R.;
RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL, BC054552; AAHS4552.1; -.
KW Hypothetical protein.
SQ SEQUENCE 545 AA; 59679 MW; 88DD35FF30DF3091 CRC64;

Alignment Scores:
Pred. No.: 6.59e-73 Length: 545
Score: 1073.00 Matches: 215
Percent Similarity: 72.58% Conservative: 47
Best Local Similarity: 59.58% Mismatches: 59
Query Match: 34.50% Indels: 40
DB: 11 Gaps: 5

US-09-270-437D-5 (1-1708) x Q7TQF9 (1-545)
Qy 38 GCATCATGAAGCTCAATGGCCACAGTTGGAGACCATGCCTGAAGTCTCCTACATC 97
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 135 AlaIleGluLysLeuSerGlyHisGlnPheGluAspTyrSerPheLysIleSerTyrIle 154
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 98 CCGGATGACGAGATAGCA-----CAGGAGCTCAGATGGCGCGGAGGGGCTTGGC 151
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 155 ProAspGluGluValSerSerProSerProHisArgAlaArgGluGlnGlyHisGly 174
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 152 TCTCGGGTTCAGCCCGCAGGGCTCACTGTGGCAGCGGGGGCCCCAGCCAGCAGCAG 211
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 175 -----ProGlySerSerSerGlnAlaArg 182
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 212 CAAGTGGACATCCCTCTCGGCTCTGTGCCACCCAGTATGTGGTGCCTATTATGSC 271
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 183 GlnIleAspPheProLeuArgIleLeuValProThrGlnPheValGlyAlaIleIleGly 202
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 272 AAGGAGGGGCCACCATCCGCAACATCAAAACAGACCCAGTCCAGATAGACGTGCAT 331
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 203 LysGluGlyLeuThrIleLysAsnIleThrLysGlnThrGlnSerArgValAspIleHis 222
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 332 AGAAGGAGAGCGCAGGTGCAGCTGAAGAGCCATCAGTGTGCATCCACCCCTCAGGCG 391
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 223 ArgLysGluAsnSerGlyAlaAlaGlnLysProValThrIleHisAlaThrProGluGly 242
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 392 TGTCTCTCGCTTGTAAAGATGATCTTGGAGATTATGCATAAAGAGCGCTAAAGACACCAA 451
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 243 ThrSerGluAlaCysArgMetIleLeuGluIleMetGlnLysGluAlaAspGluThrLys 262
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 452 AGCGTGTGAGGTTCCCTCTGAAGATCTGGCCCTTAATACTTTGTAGGGGCTTCATT 511
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 263 LeuAlaGluGluValProLeuLysIleLeuAlaHisAsnGlyPheValGlyArgLeuIle 282
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 512 GCGAAGGAGGACGGAACCTGAAGAGGTAGAGCAAGATACCGAGACAAAATACCATC 571
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 283 GlyLysGluGlyArgAsnLeuLysLysIleGluHisGluThrGlyThrLysIleThrIle 302
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 572 TCTCGTTGCAAGACTTACCTTTACACCTGTGAGAGACCATCCTGTGAAGGGGGCC 631
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Db 303 SerSerLeuGlnAspLeuSerIleTyrAsnProGluArgThrIleThrValArgGlyThr 322
 QY 632 ATCCAGATTGTTCCAGCGCCGACGAGAAATATGAAAGTTCGGGAGGCTATGAG 691
 Db 323 IledGuaLacYsAlaAsnAlaGluileGluileMetLysLysLeuArgGluAlaPheGlu 342
 QY 692 AATGATGTGGCTGCATGAGC-----TCTCACTGATGCTCCCTGCGCTGACCTGGCTGCT 745
 Db 343 AsnAspMetLeuAlaValAsnGlnInAlaAsnLeuileProGlyLeuAsnLeuSerAla 362
 QY 746 GTAGTGCTTTTCCAGCTTATCCAGCGCAGTCCCGCGCT-----CCACACACGCTT 799
 Db 363 LeuGlyIlePheSerThrGlyLeuSerValLeuProProAlaGlyProArgGlyVal 382
 QY 800 ACTGGGCTGCTCCCTATAGCTCCTTTATG-----829
 Db 383 ProProSerProProTyrHisProPheAlaThrHisSerGlyTyrPheSerSerLeuTyr 402
 QY 830 -----CAGCTCCGAGCAGGAGTGGTG 853
 Db 403 ProHisHisPheGlyProPheProHisHisSerTyrProGluGlnGluThrVal 422
 QY 854 CAGGTGTTTATCCCGCCGAGGAGTGGCGCATCATCGCAAGAGGGGACGACATC 913
 Db 423 SerLeuPheIleProThrGlnAlaValGlyAlaIleIleGlyLysLysGlyAlaHisIle 442
 QY 914 AAACAGCTCTCCCGTTGTCAGCGCTCCATCAAGATTGCACACCCGAAACACCTGAC 973
 Db 443 LysGlnLeuAlaArgPheAlaGlyAlaSerIleLysIleAlaProAlaGluGlyProAsp 462
 QY 974 TCCAAAGTCGTGTATCATCATCAGTGGACCGCCAGAGCCCAATTCAGGCTCAGGGA 1033
 Db 463 ValSerGluArgMetValIleIleThrGlyProProGluAlaGlnPheLysValSerAla 482
 QY 1034 AGA 1036
 Db 483 Lys 483
 RESULT 15
 Q9VZ69 PRELIMINARY; PRT; 566 AA.
 AC Q9VZ69; Q9NSG9;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE CG1691 protein (IGF-II mRNA-binding protein) (SD07045p).
 GN IMP OR CG1691.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkeley;
 RX MEDLINE=20196006; PubMed=107311132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Anantides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abrell J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J.J., Brokstein P., Brothier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hoslin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Staden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
 RA Evans C.A., Gocayne J.D., Anantides P.G., Brandon R.C., Rogers Y.,
 RA Banson J., An H., Baldwin D., Banton J., Beeson K.Y., Busam D.A.,
 RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
 RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
 RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,
 RA Gonzalez C., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
 RA Ibegwam C., Jalali M., Kruse D., Li P., Martel B., Moshrefi A.,
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
 RA Pacleb J., Paragass V., Park S., Patel S., Pfeiffer B.,
 RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
 RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
 RT "Sequencing of Drosophila melanogaster genome.";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
 RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,
 RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
 RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
 RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
 RA Searle S.W.J., Smith E., Shu S., Smutniak F., Whitfield E., Lewis S.E.,
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
 RT "Annotation of Drosophila melanogaster genome.";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RA FlyBase;
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 RN [6]
 RP SEQUENCE FROM N.A.
 RA Nielsen J., Nielsen F.C., Christiansen J.;
 RT "Cloning and Expression of a Drosophila Homolog of IMP/Vgl-RBP.";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [7]
 RP SEQUENCE FROM N.A.
 RA Stapleton M., Brokstein P., Hong L., Agbavani A., Carlson J.,
 RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
 RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
 RA Nunoo J., Pacleb J., Paragass V., Park S., Phouanavong S., Wan K.,
 RA Yu C., Lewis S.E., Rubin G.M., Celniker S.;
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 CC -!- SIMILARITY: CONTAINS 4 KH DOMAINS.
 DR EMBL; AE003484; AAF47958.2; -.
 DR EMBL; AF241237; AAF63331.1; -.

DR EMBL; AY069821; AAL39966.1; --
 DR FlyBase; FBgn0030235; Imp.
 DR GO; GO:0003676; F.nucleic acid binding; IEA.
 DR InterPro; IPR004087; KH_dom.
 DR InterPro; IPR004088; KH_type_1.
 DR Pfam; PF00013; KH; 4.
 DR SMART; SM00322; KH; 4.
 DR PROSITE; PS00084; KH_TYPE 1; 4.
 DR SEQUENCE 566 AA; 52129 MW; B4C74C5C99B0C830 CRC64;

Alignment Scores:
 Pred. No.: 1,33e-56 Length: 566
 Score: 858.50 Matches: 200
 Percent Similarity: 57.43% Conservative: 86
 Best Local Similarity: 40.16% Mismatches: 150
 Query Match: 27.60% Indels: 62
 DB: Gaps: 11

US-09-270-437D-5 (1-1708) x Q9VZ69 (1-566)

QY 26 TTATCCCGGGAGCCCATCATGAAGCTGAATGGCCACCAAGTTGGAGAACCATCCCTGAAG 85
 Db 29 LeuAspArgAlaAlaVal---GlyLeuAsnGlyValGluPheGluGlySerLysLeuHis 47
 QY 86 GTCTCTACATCCCGATGAGCAGATAGCACAGGACCTGAGNAATGGCGCCGAGGGGC 145
 Db 48 AlaGluGlnLeuAspLysAsnGln----- 55
 QY 146 TTTGGCTCTCGGGTCTCAGCCCGCCAGGGCTCACCTGTGTCGACGGGGGCCCGACCAAG 205
 Db 56 -----ArgArgSerGlnArgAsnGlnArgAsnProTyr---ProGlyMetProGlyPro 72
 QY 206 CACAGCAAGTGAACATCCCTTCCTGGTCCGCCACCCAGTATGTGGTGCCATT 265
 Db 73 GlyArgGlnAlaAspPheProLeuArgIleLeuValGlnSerGluMetValGlyAlaIle 92
 QY 266 ATTGGCAAGAGGGGCCACCATCCGCAACATCACAAAACAGACCCAGTCCAGATAGAC 325
 Db 93 IleGlyArgGlnGlySerThrIleArgThrIleThrGlnGlnSerArgAlaArgValAsp 112
 QY 326 GTCCATAGGAAGAGAACGCGAGTGCAGCTGGAAGAACCCATCATGTGCACTCCACCCCT 385
 Db 113 ValHisArgLysGluAsnValGlySerLeuGluLysSerIleThrIleTyrGlyAsnPro 132
 QY 386 GAGGGCTGCTCCCTCCCTGTAGATGATCTTCGAGATTATGCATAAAGAGGCTAAGGAC 445
 Db 133 GluAsnCysThrAsnAlaCysLysArgIleLeuGluValMetGlnGlnAlaIleSer 152
 QY 446 ACCAAAACGGCTGACCGAGGTTCCTCGCATATCCTGGCCCATATAACTTTGAGGGCT 505
 Db 153 ThrAsnLysGly---GluIleCysLeuLysIleLeuAlaHisAsnAsnLeuIleGlyArg 171
 QY 506 CTCATTGGCAAGAGACCGAACCTGAAGAAGGTAGACAGATACCGAGACAAAATC 565
 Db 172 IleIleGlyLysSerGlyAsnThrIleLysArgIleMetGlnAspThrAspThrLysIle 191
 QY 566 ACCATCTCCTCGTTGCAAGACCTTACCCTTTACACCTGAGAGGCCATCACTGTGAAG 625
 Db 192 ThrValSerSerIleAsnAspIleAsnSerPheAsnLeuGluArgIleIleThrValLys 211
 QY 626 GGGGCCATCAGAAATTGTTCCAGGGCCGACAGCAAAATATGAAGAAAGTCCGGAGGCC 685
 Db 212 GlyLeuIleGluAsnMetSerArgAlaGluAsnGlnIleSerThrLysLeuArgGlnSer 231
 QY 686 TATCAGATGATGGGTGGCATGAGC-----TCTCCTGATCCCTGGCTGGAACCTG 739
 Db 232 TyrGluAsnAspLeuGlnAlaMetAlaProGlnSerLeuMetPheProGlyLeuHisPro 251
 QY 740 GCTGCTAGGTCTTTTCCAGCTTCA-----TCCAGCGAGTCCCGCGCCT 787
 Db 252 MetAlaMetSerThrProGlyAsnGlyMetValPheAsnThrSerMetProPhePro 271
 QY 788 -----CCACGACGGTTACTGGGGGTGCTCCCTATAGTCCCTTTATGCGAG 832

Search completed: July 13, 2004, 12:12:37
 Job time : 113 secs

Db 272 SerCysGlnSerPheAlaMetSerLysThrProAlaSerValValProValPhePro 291
 QY 833 CTTCCCGAGCAGAGATGGTCCAGTGTGTTATCCCGCCCGCAGAGTGGCGCCCATCATC 892
 Db 292 AsnAspLeuGlnGluThrThrTyrLeuTyrIleProAsnAsnAlaValGlyAlaIleIle 311
 QY 893 GGCAGAAGAGGGCCAGCACATCAACAGCTCTCCCGGTTTGCAGCGCTCCATCAAGATT 952
 Db 312 GlyThrArgGlySerHisIleArgSerIleMetArgPheSerAsnAlaSerLeuLysIle 331
 QY 953 GCACACCCGAAACA-----CCTGACTCCAAAGTTCGTATGTTGTTATCATCACT 1000
 Db 332 AlaProLeuAspAlaAspLysProLeuAspGlnGlnThrGluArgLysValThrIleVal 351
 QY 1001 GCACCGCCAGAGGCCCAATTCAAGCTCAGGAAAGATCTATGGCAAACTCAAGAGGAG 1060
 Db 352 GlyThrProGluGlyGlnTrpLysAlaGlnTyrMetIlePheGluLysMetArgGluGlu 371
 QY 1061 AACTTTTGTGTCACAGGAGGAAGTGAAGCTGAGACCCACATACGTGTGCCAGCATCA 1120
 Db 372 GlyPheMetCysGlyThrAspValArgLeuThrValGluLeuValAlaSerSer 391
 QY 1121 GCAGCTGCGCGGTCATTGGCAAGGTGGAACCGGTGAACGAGTTCGAGAATTTCGACG 1180
 Db 392 GlnValGlyArgIleIleGlyLysGlyGlnAsnValArgGluLeuGlnArgValThr 411
 QY 1181 GCAGCTGAGGTGGTAGTACCAAGA-----GACCAG 1210
 Db 412 GlySerValIleLysLeuProGluHisAlaLeuAlaProProSerGlyGlyAspGluGlu 431
 QY 1211 ACCCTGTATGAGACGACCGAGTTCATCGTGAATCATCGACATTTCTATGCCAGTCAG 1270
 Db 432 ThrPro-----ValHisIleIleGlyLeuPheTyrSerValGln 444
 QY 1271 ATGGCTCAACGAAGATCCGACATCCTGGCCCCAG----- 1306
 Db 445 SerAlaGlnArgArgIleArgAlaMetMetLeuSerThrAsnProProProIleThrLys 464
 QY 1307 -----GTTAAGCAGCAGCATCAGAAGGGACAGAGTAACCGAGGCCCGCCAGCAGG 1354
 Db 465 LysGlnLysAlaAlaLysGluGlnGlnGlnGlnSerLeuAlaGlyAla-Al 484
 QY 1355 AGGAAGTGACCGCCCTCCCTGTCCTTNGAGTCCAGGACAAACACGGCA 1406
 Db 484 aSerSerGlySerGlnGlnGlnProGlnSerProSerGlnGlnAla 501

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 13, 2004, 17:42:43 ; Search time 4592 Seconds
(without alignments)
11107.274 Million cell updates/sec

Title: US-09-270-437D-5

Perfect score: 1708

Sequence: 1 agggagcgtgcgcaccgcc.....attcttcagggttttaaaa 1708

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*

2: em_esthum:*

3: em_estlin:*

4: em_estmva:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_estc:*

9: gb_esti:*

10: gb_est2:*

11: gb_est3:*

12: gb_est4:*

13: gb_est5:*

14: gb_estfun:*

15: em_estom:*

16: em_gss_hum:*

17: em_gss_inv:*

18: em_gss_pln:*

19: em_gss_vrt:*

20: em_gss_fun:*

21: em_gss_mam:*

22: em_gss_mus:*

23: em_gss_pro:*

24: em_gss_rod:*

25: em_gss_pbg:*

26: em_gss_vrl:*

27: gb_gss1:*

28: gb_gss2:*

29: gb_gss3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1164.8	68.2	2780	11 AK013940	AK013940 Mus muscu
2	1163.2	68.1	2718	11 AK044850	AK044850 Mus muscu
3	960.6	56.2	1085	13 BUI90334	BUI90334 AGENCOURT
4	912.6	53.4	1042	12 BM928196	BM928196 AGENCOURT

5	828.4	48.5	1042	12 BM561057	BM561057 AGENCOURT
6	736.6	43.1	950	13 BQ648383	BQ648383 AGENCOURT
7	671.8	39.3	831	12 BG115319	BG115319 602316274
8	670.4	39.3	1201	13 BX401619	BX401619 BX401619
9	669	39.2	712	13 BQ225582	BQ225582 AGENCOURT
10	662	38.8	896	10 BF984962	BF984962 602308484
11	660.2	38.7	891	13 BX372273	BX372273 BX372273
12	653.4	38.3	955	13 BQ651665	BQ651665 AGENCOURT
13	647.2	37.9	764	12 BM006191	BM006191 603614013
14	627.6	36.7	983	13 BQ643920	BQ643920 AGENCOURT
15	627	36.7	4100	11 AK088465	AK088465 Mus muscu
16	625.4	36.6	2202	11 AK011689	AK011689 Mus muscu
17	586.4	34.4	978	12 BG115593	BG115593 602317291
18	582.2	34.1	1173	12 BM467135	BM467135 AGENCOURT
19	574.6	33.6	919	13 BX327672	BX327672 BX327672
20	547.6	32.1	884	13 BUI98549	BUI98549 AGENCOURT
21	546	32.0	568	14 CB146278	CB146278 K-EST0201
22	540	31.6	545	13 BX089485	BX089485 BX089485
23	537	31.4	1248	29 AY417413	AY417413 Mus muscu
24	533.8	31.3	1061	13 BQ647561	BQ647561 AGENCOURT
25	513.8	30.1	953	13 BQ647360	BQ647360 AGENCOURT
26	513.6	30.1	658	13 BQ256493	BQ256493 NISC KO03
27	513.4	30.1	656	10 AW303505	AW303505 xv19h09.x
28	512.4	30.0	550	9 AA978341	AA978341 oq40a04.s
29	498.4	29.2	643	14 CF894671	CF894671 A0138810-
30	497.2	29.1	517	10 BE550506	BE550506 7a27f01.x
31	493.8	28.9	1321	29 AY417411	AY417411 Homo sapi
32	493.4	28.9	1321	29 AY417412	AY417412 Pan trogl
33	493	28.9	3592	11 AK049196	AK049196 Mus muscu
34	492	28.8	2964	11 AK044984	AK044984 Mus muscu
35	485.4	28.4	761	9 AW003366	AW003366 wq65h07.x
36	482.6	28.3	3262	11 AK077404	AK077404 Mus muscu
37	469	27.5	484	9 AI337147	AI337147 qx83c08.x
38	459.4	26.9	486	9 AL704124	AL704124 DKP2p686B
39	458	26.8	478	9 AA196977	AA196977 zq60h02.s
40	457	26.8	628	14 CF173461	CF173461 B0924A03-
41	455.4	26.7	599	14 CF914870	CF914870 B0970D11-
42	455.4	26.7	628	14 CF172610	CF172610 B0910C03-
43	453.8	26.6	619	14 CF172952	CF172952 B0916D09-
44	453.8	26.6	628	14 CF174626	CF174626 B0941302-
45	450.4	26.4	540	14 CA561264	CA561264 K0284F02-

ALIGNMENTS

RESULT 1

AK013940

LOCUS

DEFINITION

AK013940

Mus musculus 13 days embryo head cDNA, RIKEN full-length enriched

library, clone:3110001D23 product:insulin-like growth factor 2,

binding protein 1, full insert sequence.

ACCESSION

AK013940.1 GI:12851513

VERSION

HTC; CAP trapper.

KEYWORDS

Mus musculus (house mouse)

SOURCE

Mus musculus

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Muridae; Murinae; Mus.

REFERENCE

1

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,

Itoh, M., Komno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

High-efficiency full-length cDNA cloning

JOURNAL

Meth. Enzymol. 303, 19-44 (1999)

MEDLINE

99279253

PUBMED

10349636

REFERENCE

2

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,

Itoh, M., Komno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new genes

Genome Res. 10 (10), 1617-1630 (2000)

Genome Res. 10 (10), 1617-1630 (2000)

20499374

PUBMED

11042159

REFERENCE
AUTHORS

Shibata,K., Itoh,M., Aizawa,K., Nagacka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsuunai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,C., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Wataniki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuda,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system-384-Format sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913
11076861

JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)

TITLE

The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.

JOURNAL
REFERENCE
AUTHORS

Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
6 (bases 1 to 2780)

JOURNAL
REFERENCE
AUTHORS

Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Azai,A., Aono,H., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Horii,F., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T., Kato,H., Kawai,J., Koijima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nisha,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Tejima,I., Toya,T., Yamamura,T., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
Direct Submission
Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gscl.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
Please visit our web site (<http://genome.gsc.riken.go.jp/>) for further details.
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5'-GAGAGAGAGATCCAGACTCTTTTTCCTTTTTTTTN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5'-GAGAGAGATCTCGATTAAATAAATTAACCCCCCCCCC 3']. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI.
Host: SstI.

FEATURES
source

Location/Qualifiers
1..2780
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM DB:3110001D23"
/db_xref="MG1:1904502"
/db_xref="taxon:10090"
/clone="3110001D23"
/tissue_type="head"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="13 days embryo"
279..2012
/note="unnamed protein product; insulin-like growth factor"

CDS

acknowledge.

Please visit our web site for further details.

URL: <http://genome.gsc.riken.go.jp/>

URL: <http://fantom.gsc.riken.go.jp/>

FEATURES

source

Location/Qualifiers

1..2718

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL/6J"

/db_xref="FANTOM DB:B130008A04"

/db_xref="MGI:2410206"

/db_xref="taxon:10090"

/clone="B130008A04"

/tissue_types="parthenogenote"

/clone_lib="RIKEN full-length enriched mouse cDNA library"

/dev_stage="9.5 days embryo"

217..1950

/notes="unnamed protein product; insulin-like growth factor

2, binding protein 1 (MGI:1890357)

putative"

/codon_start=1

/protein_id="BAC32119.1"

/db_xref="GI:26336871"

/translations="MNKLYIGNLNEVTPADLEKVPFAEHKISYSGOFLVKSGYAFVDC

PDEHMAKAITETSGVELQKLEIEHSVPKORSRKIQIRNIIPOLWEVLDSILA

OYGTVECEOVNTSESTAVNVTVSYNREOTROALMKLNGHOLENHALKYVIPDQIT

QGPENRGFGSRGPRGSPVAGAPAKQPDVPLRLVPTQTVGAILIGESATI

RNIYKQTSKIDVHRKENGAEKAISVHSTPEGSSACKMILEIMHKEAKDTADE

VPLKLAHNFFVRLGKGRNLKVEQDTEKTISSLDLITLYNPERTITVKAIE

NCCRAQEMKKRYREAYENDVAMSLQSHLIPOLNLAUFLFPASSAVPPPSVVG

AAPYSFPMQAEQGVQVTFPAQVGLIIGKQHIKQLSRPASASIKIAPPETPSK

RVWVITGPAAQFKAQGLYKLENFEPKVEVLETHIRVPASAAAGRVIGKX

TVNELQNLTPAEVVVPRDQTPDENDQVIKLIIGHFVSAQMRKINDILAAVQKQHQK

QSNLAQARK"

ORIGIN

Query Match 68.1%; Score 1163.2; DB 11; Length 2718;
 Best Local Similarity 82.7%; Pred. No. 2.2e-288;
 Matches 1411; Conservative 0; Mismatches 263; Indels 33; Gaps 6;

QY 30 CCGGGAGCCATCATGAAGCTGAATGGCCACCAAGTTGGAGAACCATGCGCTGAAGTCT 89
 DB 611 CCAGGCAAGCTATCATGAAGCTAAATGGCCATCACTGGAGAACCATGCGCTGAAGTCT 670

QY 90 CTTACATCCCGGTAGCAGATAGCAGGACCTGAGATGGCGCCGAGGGGGCTTTG 149
 DB 671 CTTACATACCTGATGAGCAGATTAACGAGGCTCTGAGATGGCGCTCTGGAGGCTTTG 730

QY 150 GCTCTCGGGTACGCCCGGCAGGGCTACCTGTGGCAGCGGGGGCCCGAGCAAGCAGC 209
 DB 731 GGTCTCGGGCCAGCCCGGCAAGGTCGCCCGTGGCAGCAGGGGCTCCAGCAAGCAGC 790

QY 210 AGCAAGTGGACATCCCGCTCTCGGCTCTCTGGTCCACCCAGATGTGGGTGCCATTATTG 269
 DB 791 AGCCAGTGGACATCCCTCTCGGCTCTCTGGTCCACCCAGATGTGGGTGCCATTATTG 850

QY 270 GCAAGAGGGGGCCACATCCGCAACATCAAAAACAGACCCAGTCCAAAGATAGACGTGC 329
 DB 851 GCAAGAGGGGTGCACATCCGCAACATCAAAAACAGACCCAGTCCAAAGATAGACGTGC 910

QY 330 ATAGGAGGAGGACGAGGTGAGCTGAAGAGCCATCAGTGTGCATCCACCCCTGAGG 389
 DB 911 ATAGGAGGAGGATGCGGGCGCTGCGGAGAGGCCATCAGCTGTGATTAACCCCTGAAG 970

QY 390 GCTGCTCTCCGCTTGAAGATGATCTTGGAGATATTGATTAAGAGGCTAAGGACACCA 449
 DB 971 GCTGCTCTCCGCTTGAAGATGATCTTGGAGATATTGATTAAGAGGCTAAGGACACCA 1030

QY 450 AAACGGCTGACAGAGTTCCTGAGATCTCTGGCCCAATAAATCTTTAGGGGCTCTCA 509
 DB 1031 AAACGGCAGATGAAGTTCCCTGAGATCTCTGGCTCATAACTTCGTCGGGGCACTCA 1090

QY 510 TTGGCAAGGAAGGACGGAACCTGAAGAGGTAGAGCAAGATACCAGAGCAAAAATCACCA 569

DB 1091 TTGGCAAGGAAGGGCGGAACCTGAAGAAGGTGGAGCAGGACACAGAGACGGAAGATCACCA 1150
 QY 570 TCTCTCTCGTTCAGAGACCTTACCTTTACACCTCGTGAAGACCATCCTGTGTGAAGGGG 629
 DB 1151 TCTCATCGCTCCAGACCTCAGCTCTATTAACCTCGAGAGACCATCCTGTGTGAAGGGCG 1210

QY 630 CCATCAGAAATTTGTGAGGGCCGAGCAGGAGAAATAATGAAGAAAGTTTCGGAGGCTTATG 689
 DB 1211 CCAATTGAGAACTGTTCAGGGCCGAGCAGGAGATCATGAAGAAAGTTTCGAGAGGCTTACG 1270

QY 690 AGAATGATGTGGCTGCCATGAGC-----TCTACCTGATCCCTGGGCTGAACCTGGCTG 743
 DB 1271 AGAAGAGCTGGCGCCCATGAGTTGAGTCCCACTCATCCCTGGGCTTAACCTGGCTG 1330

QY 744 CTGTAGGTCTTTTCCAGCTTCATCCAGCGCAGTCCCGCGCCCTCCAGCAGAGGTTTACTG 803
 DB 1331 CTGTAGGTCTCTTCCAGCTTCATCCAGCGCTGTCCTCTCTCTCCAGCAGTGTCAACG 1390

QY 804 GGGCTGCTCCCTATAGTCTCTTTATGAGAGCTCCGAGCAGGAGATGGTGCAGGTGTTTA 863
 DB 1391 GGGCTGCTCCCTATAGTCTCTTTATGAGAGCTCCGAGCAGGAGATGGTGCAGGTGTTTA 1450

QY 864 TCCCGCCAGGAGTGGGGCCCATCATCGCAAGAGAGGGGCGAGCAGCATCAAAACAGCTCT 923
 DB 1451 TCCCGCCAGGCTGTGGGCGCCCATCATCGCAAGAGAGGGGCGAGCAGCATCAAAACAGCTCT 1510

QY 924 CCCGGTTTGCAGCGCTCCATCAAGATTGCACACCCGAAACACCTGACTCCAAAGTTC 983
 DB 1511 CCCGGTTTGCAGCGCTCCATCAAGATTGCACACCCGAAACACCTGACTCCAAAGTTC 1570

QY 984 GTATGGTTATCATCACTGGACCGCCAGCCCAATTCAGAGCTCAGGGAGAGATCTATG 1043
 DB 1571 GAATGGTGTCTATCACTGGACCGCCAGAGCTCAGTTCAAGGGCCAGGGAGAGATCTATG 1630

QY 1044 GCAAACTCAAGAGGAGAACTTTCTTTGGTCCCAAGAGGAGAGTGAAGTGGAGACCCACA 1103
 DB 1631 GCAAACTCAAGAGGAGAACTTTCTTTGGTCCCAAGAGGAGAGTGAAGTGGAGACCCACA 1590

QY 1104 TACGTGTCCAGCATCAGCAGCTGGCCGGGTCTATGGCAAGAGTGGNAAGACCGTGAAG 1163
 DB 1691 TACGGGTTCGGGTCTCAGCAGCGCGGTCTCATCGCAAGGCGGCAAAACCGTGAATG 1750

QY 1164 AGTTGCAAGAAATTTGACCGCAGCTGAGGTGGTAGTACCAAGAGACCAAGCCCTGATGAGA 1223
 DB 1751 AGTGTGCAAGAAATTTGACCGCAGCTGAGGTGGTAGTACCAAGAGACCAAGCCCTGATGAGA 1810

QY 1224 ACGACCAAGTCTCGTGAATAATCATCGCAATTTCTATGTCAGTCAAGTGGCTCAACGGA 1283
 DB 1811 ACGACCAAGTCTCGTGAATAATCATCGCAATTTCTATGTCAGTCAAGTGGCTCAACGGA 1870

QY 1284 AGATCCAGACATCCTGGCCAGGTTAAGCAGCAGATCAGAGGGACAGTAACACGAG 1343
 DB 1871 AGATCCAGACATCCTGGCTCAAGTTAAGCAACAGCAAGAGAGGACAGCAACCTGG 1930

QY 1344 CCCAGGCAACGAGGAAGTGA-CCAGGCCCTCCCTGTCCCTTNGAGTCCAGGACAAACAAAG 1402
 DB 1931 CCCAGGCTCGAGGAAGTGAACCCCGCCCTCTCTGTCCCATTTGGCTCCAAGATCAGCAGG 1990

QY 1403 GGCAGAA-----ATCGAGAGTGTGCTCTCCCGGCGGCGCTGAG 1441
 DB 1991 AGGAACACAGAACTGGAGGGGCGGGTGGAGGGCGGTGTGCTCTTCCAGCAGGCTGAG 2050

QY 1442 AATGAGTGGGAATCCGGGACACNTGGCCGGGTGTAGATCAGGTTTGGCCACTTGATTG 1501
 DB 2051 AATGAGTGGGAATCAGGG-CATTTGGGCTGGCTGGAGATCAGGTTTGCACACTGTCTTG 2109

QY 1502 AGAAGATGTTCCAGTGAGGAACCTGTATCTNTACGCCCAACACCCACCCAAATTTGGCC 1561
 DB 2110 AGAACAATGTTCCAGTGAGGAATCTCTGCTCTCTGCCCCCA--TTGAGCAGCTGGCCA 2167

QY 1562 CAACACTGTNTGCCCTCGGGGTGTGAGAAATTTAGCGCAAGGCACTTTTAAACCTGGA 1621

```

Db      2168 CAGCCACCCCTTGGAAATATCACCAATTGCAATCATAGCTTGGGTGCTTTAAACGTGGA 2227
Qy      1622 TTGTTTAAAGAGCTCTCCAGGCCCCCACCAGAGGGTGGATCACACTCAGTCGGGAAGAA 1681
Db      2228 TTGTCT--TGAGTCTTCCAGCTCCATGGAAGATGGGTGCAGATCCAGTGGGAGAG 2285
Qy      1682 AATAAAATTCCTTCAGGTTTAAAA 1708
Db      2286 AATAAAATTCCTTCAGGTTTATAA 2312

RESULT 3
LOCUS   BU190334
DEFINITION AGENCOURT_7186901 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6001834
5', mRNA sequence.
ACCESSION BU190334
VERSION   BU190334.1 GI:22704318
KEYWORDS EST.
SOURCE   Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1085)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: rgs@nih.gov
Tissue Procurement: CGAP (Stanford)
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1CM2269 row: b column: 11
High quality sequence stop: 671.

FEATURES
source
1..1085
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6001834"
/tissue_type="hepatocellular carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_100"
/note="Organ: liver; Vector: pOTB7; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAG(G). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-cDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH_MGC
Library."

ORIGIN
Query Match 56.28; Score 960.6; DB 13; Length 1085;
Best Local Similarity 95.08; Pred. No. 2:5e-236;
Matches 1026; Conservative 0; Mismatches 45; Indels 9; Gaps 3;

Qy      30 CCGGGGAGCCATCATGAAGCTGAAT-GGCCACAGTTGGAGAACCATGCCCTGAAGGTC 88
Db      5 CCAGGCAAGCCATCATGAAGCTGAATGGGCCACCAGTTGGAGAACCATGCCCTGAAGGTC 64
Qy      89 TCCTACATCCCGATGACGATAGACACAGGACCTGAGAAATGGGCGCCGAGGGGCTTT 148
Db      65 TCCTACATCCCGATGACGATAGACACAGGACCTGAGAAATGGGCGCCGAGGGGCTTT 124
Qy      149 GGCTCTCGGGTCAGCCCGCCAGGGCTCACCTGTGGCAGCGGGGGCCCCAGGCAAGCAG 208
Db      125 GGCTCTCGGGTCAGCCCGCCAGGGCTCACCTGTGGCAGCGGGGGCCCCAGGCAAGCAG 184

```

```

Qy      209 CAGCAAGTGGACATCCCCCTTTCGGCTCCTGGTGCCACCAGTATGTGGGTGCCATATT 268
Db      185 CAGCAAGTGGACATCCCCCTTTCGGCTCCTGGTGCCACCAGTATGTGGGTGCCATATT 244
Qy      269 GGCAGAGAGGGGGCCACCATCCGAAACATCAAAACAGACCCAGTCCAAAGATAGACGTG 328
Db      245 GGCAGAGAGGGGGCCACCATCCGAAACATCAAAACAGACCCAGTCCAAAGATAGACGTG 304
Qy      329 CATAGGAAGGAGAACGCGAGGTGCAGTGAAGAGCCATCAGTGTGCATCCACCCCTCAG 388
Db      305 CATAGGAAGGAGAACGCGAGGTGCAGTGAAGAGCCATCAGTGTGCATCCACCCCTCAG 364
Qy      389 GGCTCTCTCTCCGCTTGAAGATGATCTTGAGATTTATGCATAAAGAGGCTTAAGACACC 448
Db      365 GGCTCTCTCTCCGCTTGAAGATGATCTTGAGATTTATGCATAAAGAGGCTTAAGACACC 424
Qy      449 AAAACGGCTGACGAGGTTCCCTGGAAGATCTGGCCCATATAAATTTGTAGGGGCTCTC 508
Db      425 AAAACGGCTGACGAGGTTCCCTGGAAGATCTGGCCCATATAAATTTGTAGGGGCTCTC 484
Qy      509 ATTGCAAGGAAGGACGGAACTGGAAGAGGTAGAGCAAGATACCGAGACAAAATCACC 568
Db      485 ATTGCAAGGAAGGACGGAACTGGAAGAGGTAGAGCAAGATACCGAGACAAAATCACC 544
Qy      569 ATCTCTCTGTTGCAAGACCTTACCCCTTTACAACTGAGAGACCATCCTGTGAGGGG 628
Db      545 ATCTCTCTGTTGCAAGACCTTACCCCTTTACAACTGAGAGACCATCCTGTGAGGGG 604
Qy      629 GCATCGAGAAATTTGTCAGGGCCGAGCAGAAATAATGAAGAAAGTTCCGGAGGCTTAT 688
Db      605 GCATCGAGAAATTTGTCAGGGCCGAGCAGAAATAATGAAGAAAGTTCCGGAGGCTTAT 664
Qy      689 GAGAATGATGTGGCTGCCATGAGC-----TCTCAGCTGATCCCTGGGCTGAACTGGCT 742
Db      665 GAGAATGATGTGGCTGCCATGAGC-----TCTCAGCTGATCCCTGGGCTGAACTGGCT 724
Qy      743 GCTGTAGGTCTTTTCCAGGCTTCATCCAGCGAGTCCCGCCGCTCCAGAGAGGCTTACT 802
Db      725 GCTGTAGGTCTTTTCCAGGCTTCATCCAGCGAGTCCCGCCGCTCCAGAGAGGCTTACT 784
Qy      803 GGGGCTGTCTCTATAGTCTCTTTATGAGGCTCCGAGAGGAGATGGTGCAGGTGTTT 862
Db      785 GGGGCTGTCTCTATAGTCTCTTTATGAGGCTCCGAGAGGAGATGGTGCAGGTGTTT 844
Qy      863 ATCCCGGCGCCAGGAGTGGGGCCATCATCGGAGAGAGGGGCGAGCATCAACAGCTC 922
Db      845 ATCCCGGCGCCAGGAGTGGGGCCATCATCGGAGAGAGGGGCGAGCATCAACAGCTC 904
Qy      923 TCCCGGTTTGCAGGCGCTTCATCAAGATTGCACCCCGAAACACCTGACTCCAAAGTT 982
Db      905 TCCCGGTTTGCAGGCGCTTCCTCAAGATTGCACCCCGAAACACCTGACTCCAAAGTT 964
Qy      983 CGTATGGTTATCATCAC--TGACCGCCAGAGGCCCAATTCAAGCTCAGGAGAGAACT 1040
Db      965 CGAATGGTTATTCATACCTGGAACGCCCAAGGGCCCAATTCAGGCTCAGGAGAAATCT 1024
Qy      1041 ATGGCAAACTCAAGGAGGAGAACTTTTGTGTCACCAAGGAGAAAGTGAAGCTGGAGACC 1100
Db      1025 ATGGCAATTCAGGAGGAGAAACTCTTTGGCCCGGAGGAGATGAACCTTGGGAACC 1084

RESULT 4
LOCUS   BM928196
DEFINITION AGENCOURT_6715416 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:5797961
5', mRNA sequence.
ACCESSION BM928196
VERSION   BM928196.1 GI:19378575
KEYWORDS EST.
SOURCE   Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 1042)
 NIH-MGC <http://mgi.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: CGAP (Stanford)
 cDNA Library Preparation: Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
 Plate: LLM2263 row: c column: 18
 High quality sequence stop: 662.
 Location/Qualifiers
 1. 1042
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:579761"
 /tissue_type="hepatocellular carcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH MGC 100"
 /note="Organ: liver; Vector: pOTB7; Site: 1: XhoI; Site 2:
 EcoRI; cDNA made by oligo-dT priming. Directionally cloned
 into EcoRI/XhoI sites using the following 5' adaptor:
 GGACGAG(G). Size-selected >500bp for average insert size
 1.8kb. Library constructed by Ling Hong in the laboratory
 of Gerald M. Rubin (University of California, Berkeley)
 using ZAP-cDNA synthesis kit (Stratagene) and Superscript
 II RT (Life Technologies). Note: this is a NIH_MGC
 Library."

Query Match 53.4%; Score 912.6; DB 12; Length 1042;
 Best Local Similarity 97.5%; Pred. No. 5.9e-224;
 Matches 981; Conservative 0; Mismatches 15; Indels 10; Gaps 5;

QY 30 CCGGGGAGCCATCATGAAGCTCAATGAGCCAGGTTGGAGACCATGCTCCCTGAGGCT 89
 DB 5 CCAGGCGAGCCATCATGAAGCTCAATGAGCCAGGTTGGAGACCATGCTCCCTGAGGCT 64
 QY 90 CTTACATCCCCGATGAGCAGATAGCAGGACCTGAGAAATGGGCCCGGAGGGGGCTTTG 149
 DB 65 CTTACATCCCCGATGAGCAGATAGCAGGACCTGAGAAATGGGCCCGGAGGGGGCTTTG 124
 QY 150 GCTCTGGGGTACGCCCCCGCCAGGCTCCTGAGGACCTGAGAAATGGGCCCGGAGGGGGCTTTG 209
 DB 125 GCTCTGGGGTACGCCCCCGCCAGGCTCCTGAGGACCTGAGAAATGGGCCCGGAGGGGGCTTTG 184
 QY 210 AGCAAGTGGACATCCCCCTTCGCTCCTGAGGACCTGAGAAATGGGCCCGGAGGGGGCTTTG 269
 DB 185 AGCAAGTGGACATCCCCCTTCGCTCCTGAGGACCTGAGAAATGGGCCCGGAGGGGGCTTTG 244
 QY 270 GCAAGAGGGGGCCATCCGCAACATCAGAAACAGCCAGGTTGAGGACCTGAGGACCTGAGG 329
 DB 245 GCAAGAGGGGGCCATCCGCAACATCAGAAACAGCCAGGTTGAGGACCTGAGGACCTGAGG 304
 QY 330 ATAGGAAGGAGAACGCGAGGTGAGCTGAGAAAGCCATGAGTGGCACTCCACCCCTGAGG 389
 DB 305 ATAGGAAGGAGAACGCGAGGTGAGCTGAGAAAGCCATGAGTGGCACTCCACCCCTGAGG 364
 QY 390 GCTGCTCCTCGCTTGTAGATGATCTTGGAGATATGATGATGATGATGATGATGATGATGAT 449
 DB 365 GCTGCTCCTCGCTTGTAGATGATCTTGGAGATATGATGATGATGATGATGATGATGATGAT 424
 QY 450 AAACGGCTGACGAGGTTCCTCGGAGATCTTGGCCCATTAATACCTTTGAGGGCGCTTCA 509
 DB 425 AAACGGCTGACGAGGTTCCTCGGAGATCTTGGCCCATTAATACCTTTGAGGGCGCTTCA 484
 QY 510 TTGGCAAGGAGGACGGAACCTTGAAGAGGTGAGGACGATACCGAGACAAATAATCAACCA 569

Db 485 TTGGCAAGGAGGACGGAACCTGAAGAGGTGAGGACGATACCGAGACAAATAATCAACCA 544
 QY 570 TCTCTCTGTTCAAGACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTT 629
 Db 545 TCTCTCTGTTCAAGACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTT 604
 QY 630 CCAATGAGATGTTGCGAGGCGGCGAGGAGGAAATAATGAAGAAAGTTTCGGAGAGGCTTATG 689
 Db 605 CCAATGAGATGTTGCGAGGCGGCGAGGAGGAAATAATGAAGAAAGTTTCGGAGAGGCTTATG 664
 QY 690 AGAATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 743
 Db 665 AGAATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 724
 QY 744 CTGTAGGCTTTTCCAGGCTTCATCCAGCGCGATCCCGCGGCTCCAGAGGCTTACTG 803
 Db 725 CTGTAGGCTTTTCCAGGCTTCATCCAGCGCGATCCCGCGGCTCCAGAGGCTTACTG 784
 QY 804 GGGCTGCTCCCTATAGCTCCCTTATGAGGCTCCCGAGGAGGAGGAGGAGGAGGAGGAGGAGG 863
 Db 785 GGGCTGCTCCCTATAGCTCCCTTATGAGGCTCCCGAGGAGGAGGAGGAGGAGGAGGAGGAGG 844
 QY 864 T-CCCCCGCCAGGAGGCTGGCGCCATCATCGGCAAGAGGCGGAGGAGGAGGAGGAGGAGGAGG 922
 Db 845 TCCCCCGCCAGGAGGCTGGCGCCATCATCGGCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 904
 QY 923 TCCCGGTTTGGCAGGCTTCATCAAGATTGCAACCCAGGAGGAGGAGGAGGAGGAGGAGGAGG 981
 Db 905 TCCCGGTTTGGCAGGCTTCATCAAGATTGCAACCCAGGAGGAGGAGGAGGAGGAGGAGGAGG 964
 QY 982 TCGTAT-GGTTTATCATCAC-TGGACCGCCAGAGGCGGCAATTCAGG 1025
 Db 965 TCGTATGGGTTATCATCCCTTGGACCGGAGAGGCCCAATTTCAAG 1010

RESULT 5
 BM561057 1042 bp mRNA linear EST 20-FEB-2002
 LOCUS AGENCOURT_6566228 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:5550887
 DEFINITION 5', mRNA sequence.
 ACCESSION BM561057
 VERSION BM561057.1 GI:18805998
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 1042)
 NIH-MGC <http://mgi.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
 Plate: LLM2263 row: 1 column: 24
 High quality sequence stop: 682.
 Location/Qualifiers
 1. 1042
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:5550887"
 /tissue_type="retinoblastoma"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH MGC 67"
 /note="Organ: eye; Vector: pCMV-SPORT6; Site 1: NotI;

FEATURES
 source

Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.75 kb. Library constructed by Life Technologies."

ORIGIN		Query Match	48.5%; Score 828.4; DB 12; Length 1042;
		Best Local Similarity	96.8%; Pred. No. 3.1e-202;
		Matches 920; Conservative 0; Mismatches 18; Indels 12; Gaps 7;	
QY	438	CTAAGGACACCAAAACGGCTGACGAGTTCCTCCCTGAAGATCCTGGCCCATATAAATTTC	497
Db	2	CTCAGGACACCAAAACGGCTGACGAGTTCCTCCCTGAAGATCCTGGCCCATATAAATTTC	61
QY	498	TAGGGCGTCTCATTTGCCAAGGAGGACGGAACCTGAAGAGGTAGAGCAAGATACCGAGA	557
Db	62	TAGGGCGTCTCATTTGCCAAGGAGGACGGAACCTGAAGAGGTAGAGCAAGATACCGAGA	121
QY	558	CAAAATACCATCTCTCTGTTCCAAAGACCTTACCTTTACAACTCTGAGAGACCATCA	617
Db	122	CAAAATACCATCTCTCTGTTCCAAAGACCTTACCTTTACAACTCTGAGAGACCATCA	181
QY	618	CTGTGAAGGGGGCATCGAGAATTTTGCAGGGCCGAGGAGGAAATTAATGAAGAAAGTTC	677
Db	182	CTGTGAAGGGGGCATCGAGAATTTTGTTCAGGGCCGAGGAGGAAATTAATGAAGAAAGTTC	241
QY	678	GGGAGGCGCTATGAGATGATGGCTGCCATGAGC-----TCTCACTGATCCCTGGCC	731
Db	242	GGGAGGCGCTATGAGATGATGGCTGCCATGAGCCTGCGATCTCACTGATCCCTGGCC	301
QY	732	TGAACCTGGCTGTAGTGTCTTTTCCAGCTTCATCCAGCGCAGTCCCGCCCTCCCA	791
Db	302	TGAACCTGGCTGTAGTGTCTTTTCCAGCTTCATCCAGCGCAGTCCCGCCCTCCCA	361
QY	792	GCAGCGTTACTGGGCTGCTCCCTATAGTCTCTTTATGCAAGTTCCTCCGAGCAGGATGG	851
Db	362	GCAGCGTTACTGGGCTGCTCCCTATAGTCTCTTTATGCAAGTTCCTCCGAGCAGGATGG	421
QY	852	TGCAGGTGTTTATCCCGCCGACGAGTGGCGGCCCATCATCGGCAAGAGGGGACAGACA	911
Db	422	TGCAGGTGTTTATCCCGCCGACGAGTGGCGGCCCATCATCGGCAAGAGGGGACAGACA	481
QY	912	TCAACAGCTCTCCGGTTTGCAGCGCTCCATCAAGATTGACACACCGAACAACCTG	971
Db	482	TCAACAGCTCTCCGGTTTGCAGCGCTCCATCAAGATTGACACACCGAACAACCTG	541
QY	972	ACTCCAAAGTTGTTATGTTTATCATCTGACCGCGCAGAGGCCCAATTCAAGGCTCAGG	1031
Db	542	ACTCCAAAGTTGTTATGTTTATCATCTGACCGCGCAGAGGCCCAATTCAAGGCTCAGG	601
QY	1032	GAAGAATCTATGGCAAACTCAAGGAGGAGAACTTCTTTGTCCCAAGGAGGAGTGAAGC	1091
Db	602	GAAGAATCTATGGCAAACTCAAGGAGGAGAACTTCTTTGTCCCAAGGAGGAGTGAAGC	661
QY	1092	TGGAGACCCATACATGTCGACGATCAGACGCTGGCCGGTTCATTGGCAAAAGGTGGAA	1151
Db	662	TGGAGACCCATACATGTCGACGATCAGACGCTGGCCGGTTCATTGGCAAAAGGTGGAA	721
QY	1152	AAACGGTGAACGAGTTGCAGAAATTTGACGGCAGCTGAGGTGGTAGTACCAAGAGACAGA	1211
Db	722	AAACGGTGAACGAGTTGCAGAAATTTGACGGCAGCTGAGGTGGTAGTACCAAGAGACAGA	781
QY	1212	CCCTGTATGAGAACGACAGGTGATCGTGAATATCATC--GGACATTTCTATGCCAGTCA	1270
Db	782	CCCTGTATGAGAACGACAGGTGATCGTGAATATCATCGGAGCAATTTCTATGCCAGTCA	841
QY	1271	ATGGCTCAAC--GGAAGATCCGAGACATCTCT--GGCCAGGTTTAAGCAGCAGCAT--CAGAAG	1327
Db	842	ATGGCTCAACGGGAGATCCGAGACATCTCTGGGCCCGAGGTTAAGCAGCAGCATCCAAAG	901
QY	1328	GGACAGA--GTAAACAGGCCCGAGGAC--GGAGGAGATGACAGGCCCTCC	1375
Db	902	GGACAGAGGTAAACAGGCCCGAGGACGGGAGGAGATGACAGGCCCTCC	951

RESULT 6
BQ648383

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BQ648383
AGENCOURT 8342265 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6268202
5', mRNA sequence.

BQ648383
BQ648383.1 GI:21772555

EST.

Homo sapiens (human)

Homo sapiens

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 950)

NIH-MGC <http://mgs.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: CGAP (Stanford)

cDNA Library Preparation: Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

Cloning strategy: Agencourt Bioscience Corporation

Cloning strategy: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: L10M2439 row: m column: 03

High quality sequence stop: 599.

Location/Qualifiers

1. 950

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:6268202"

/tissue_type="hepatocellular carcinoma, cell line"

/lab_host="DH10B (phage-resistant)"

/clone_lib="NIH_MGC_100"

/note="Organ: liver; Vector: pOTB7; Site 1: XhoI; Site 2:

EcoRI; cDNA made by oligo-dT priming. Directionally cloned

into EcoRI/XhoI sites using the following 5' adaptor:

GGCAGGAG(G). Size-selected >500bp for average insert size

1.8kb. Library constructed by Ling Hong in the laboratory

of Gerald M. Rubin (University of California, Berkeley)

using ZAP-cDNA synthesis kit (Stratagene) and Superscript

II RT (Life Technologies). Note: this is a NIH_MGC

Library."

ORIGIN

Query Match

Best Local Similarity

Matches 857;

Conservative 0;

Mismatches 39;

Indels 17;

Gaps 8;

QY

30

CCCCGGGAGCATCATGAAGCTGAATGCCACCCAGTTCGAGAACCATGCCCTGAAGGTCT

89

Db

5

CCAGCAAGCCATCATGAAGCTGAATGCCACCCAGTTCGAGAACCATGCCCTGAAGGTCT

64

QY

90

CCTACATCCCCGATGAGCAGATAGCAGGACCTGAGAAATGGGCGCCGAGGGGCTTTG

149

Db

65

CCTACATCCCCGATGAGCAGATAGCAGGACCTGAGAAATGGGCGCCGAGGGGCTTTG

124

QY

150

GCTCTCGGGGTGAGCCCGCCAGGCTCACTGTGGCAGCGGGGCCCGCCAGCAGCAGC

209

Db

125

GCTCTCGGGGTGAGCCCGCCAGGCTCACTGTGGCAGCGGGGCCCGCCAGCAGCAGC

184

QY

210

AGCAAGTGGACATCCCCCTTCGGTCTCTGGTGCACCCAGTATGTGGTGCCTATTATG

269

Db

185

AGCAAGTGGACATCCCCCTTCGGTCTCTGGTGCACCCAGTATGTGGTGCCTATTATG

244

QY

270

GCAAGGAGGGGGCCACCATCCGCAATCAAAACAGACCCAGTCCAGATAGCGTGC

329

Db

245

GCAAGGAGGGGGCCACCATCCGCAATCAAAACAGACCCAGTCCAGATAGCGTGC

304

QY

330

ATAGGAAGGAGAACAGCAGTGTCAGCTGAAAGGCCATCAGTGTGCACTCCACCCCTGAGG

389

Db 305 ATAGGAAGGAAACGCGAGGTGCGAGCTGAAAAGCCATCAGTGGCACTCCACCCCTGAGG 364
 Qy 390 GCTCCTCCTCCGCTTGTAAAGATGATCTTGGAGATTATGCAATAAGAGGCTTAAGGACACCA 449
 Db 365 GCTCCTCCTCCGCTTGTAAAGATGATCTTGGAGATTATGCAATAAGAGGCTTAAGGACACCA 424
 Qy 450 AAACGGCTGACGAGGTTCCCTGGAAGATCTGCGCCCAATAATACTTTGTAGGGCGTCTCA 509
 Db 425 AAACGGCTGACGAGGTTCCCTGGAAGATCTGCGCCCAATAATACTTTGTAGGGCGTCTCA 484
 Qy 510 TTGCAAGGAGGACGAACTGGAAGAGGTAGGCAAGTATCGGAGACCAAAATCACCA 569
 Db 485 TTGCAAGGAGGACGAACTGGAAGAGGTAGGCAAGTATCGGAGACCAAAATCACCA 544
 Qy 570 TCTCCTCGTTGCAAGACCTTTACCTTTTACAACTGAGAGGACCATCACT-GTGAAGGGG 628
 Db 545 TCTCCTCGTTGCAAGACCTTTACCTTTTACAACTGAGAGGACCATCACTGTTGAGGGG 604
 Qy 629 GCCATCGAGAAATTGTCAGGCGCGAGCAGGAAATATGAGAAATTCGGAGGCGCTAT 688
 Db 605 GCCATCGAGAAATTGTCAGGCGCGAGCAGG-AAATAATGAAGAAATTCGGAGGCGCTAT 663
 Qy 689 GAGAATGATGTGGCTGCATGAGC-----TCTCACTGATCCTCGCCTGGAACCT--GGC 741
 Db 664 GAGAATGATGTGGCTGCATGAGCCTGCGCTCTCACTGATCCTGCGCTGAAACTGGGC 723
 Qy 742 TGCTGTAGGCTTTTCCAGCTTATCCAGCGAGTCCGCGCCTCCAGCAGC-GTTA 800
 Db 724 TGCTGTAGGCTTTTCCAGCTTATCCAGCGAGTCCGCGCCTCCAGCAGCGGTTA 783
 Qy 801 CTGGGCTGCT-CCCTATAGCTCCTTTATGAGGCTCCGAGCAGGAGATGTTG--CAGG 857
 Db 784 CTGGGCTGCTCCCTATAGCTCCTTTAGCAGCTCCCAACCGGAAAGGGGGCAAGG 843
 Qy 858 TGTTTATCCCGCCCGAGGAGTGGGCGC-----CATCATCGCAAGAGGGGCGAGCATC 913
 Db 844 GGTAAACCCCGCCCGAGGAGTGGGCGCCCATCATCGCGCAAAAGGGGCTAGCACCTC 903
 Qy 914 AAACAGCTCTCCC 926
 Db 904 AAACCGCTTCCC 916

RESULT 7
 BG115319
 LOCUS 831 bp mRNA linear EST 30-JAN-2001
 DEFINITION 602316274F1 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:4416354 5',
 mRNA sequence.
 BG115319
 VERSION BG115319.1 GI:12608825
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 831)
 NIH-MGC http://mgi.nci.nih.gov/.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgabbs@mail.nih.gov
 Tissue Procurement: ARCC
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILLNL)
 DNA Sequencing by: Inyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/ILLNL at:
 http://image.llnl.gov
 Plate: LIA10147 row: d column: 19
 High quality sequence stop: 731.
 Location/Qualifiers
 1. .831

FEATURES
 source

/organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:4416354"
 /tissue_type="duodenal adenocarcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH_MGC_88"
 /note="Organ: small intestine; Vector: pCMV-SPORT6;
 Site 1: NotI; Site 2: SalI; Cloned unidirectionally;
 oligo-dT primed. Average insert size 1.767 kb. Library
 enriched for full-length clones and constructed by Life
 Technologies. Note: this is a NIH_MGC Library."

ORIGIN

Query Match 39.3%; Score 671.8; DB 12; Length 831;
 Best Local Similarity 97.4%; Pred. No. 7e-162;
 Matches 713; Conservative 0; Mismatches 15; Indels 4; Gaps 3;
 Qy 820 CTCCTTTATGAGGCTCCGAGCAGGAGATGGTGCAGTGTATATCCCCCGCAGGCGT 879
 Db 1 CTCCTTTATGAGGCTCCGAGCAGGAGATGGTGCAGTGTATATCCCCCGCAGGCGT 60
 Qy 880 GGGCGCCATCATCGGCAAGAGGGGCGAGCATCAAAAGCTCTCCCGTTTGCAGCGC 939
 Db 61 GGGCGCCATCATCGGCAAGAGGGGCGAGCATCAAAAGCTCTCCCGTTTGCAGCGC 120
 Qy 940 CTCATCAAGATTGACACCCGAAACACCTGACTCCAAAGTTGCTATGATTCATCATC 999
 Db 121 CTCATCAAGATTGACACCCGAAACACCTGACTCCAAAGTTGCTATGATTCATCATC 180
 Qy 1000 TGGACCCCGAGAGCCCAATTCAGGCTCAGGGAAGATCTATGGCAAACTCAAGAGGA 1059
 Db 181 TGGACCCCGAGAGCCCAATTCAGGCTCAGGGAAGATCTATGGCAAACTCAAGAGGA 240
 Qy 1060 GAATCTTTTGGTCCCAAGAGGAGGAAGTGAAGCTGGAGACCCACATAGTGTGCCAGATC 1119
 Db 241 GAATCTTTTGGTCCCAAGAGGAGGAAGTGAAGCTGGAGACCCACATAGTGTGCCAGATC 300
 Qy 1120 AGCAGCTGGCGGGTCAATTCGCAAAAGTGGAAACCGTGAAACGAGTTGCAGAAATTGAC 1179
 Db 301 AGCAGCTGGCGGGTCAATTCGCAAAAGTGGAAACCGTGAAACGAGTTGCAGAAATTGAC 360
 Qy 1180 GGCAGCTGAGTGTGTAGTACCAAGAGACCAAGACCCCTGTATGAGAACGACAGGTCATCGT 1239
 Db 361 GGCAGCTGAGTGTGTAGTACCAAGAGACCAAGACCCCTGTATGAGAACGACAGGTCATCGT 420
 Qy 1240 GAAATCATCGACATTTCTATCCAGTTCAGATGGCTCAACGGAAGATCCGAGATCCT 1299
 Db 421 GAAATCATCGACATTTCTATCCAGTTCAGATGGCTCAACGGAAGATCCGAGATCCT 480
 Qy 1300 GGCCAGGTTAAGCAGCAGCATCAGAGGACAGAGTAACAGGCCCGCAGGACGAGGAA 1359
 Db 481 GGCCAGGTTAAGCAGCAGCATCAGAGGACAGAGTAACAGGCCCGCAGGACGAGGAA 539
 Qy 1360 GTGACAGCCCTCCCTGTCCTTNGAGTCCAGGACAAACGCGGAGAAATCCAGAGTG 1419
 Db 540 GTGACAGCCCTCCCTGTCCTTNGAGTCCAGGACAAACGCGGAGAAATCCAGAGTG 599
 Qy 1420 TGCTCTCCCC-GGCAGGCTGAGAAATGAGTGGGAATCCGGACACNTGGCGGGCTGTA 1478
 Db 600 TGCTCTCCCCGGGAGGCGCTGAGAAATGAGTGGGAATCCGGGACACCTGGCGGGCTGTA 659
 Qy 1479 GATCAGGTTTGGCCACTGTATTGAGAAAGATGTTCCAGTGAAGAACCTGATCTNTCAGC 1538
 Db 660 GATCAGGTTTGGCC--ACTTGATGAGAAAGATGTTCCAGTGAAGAACCTGATCTCTAGCC 717
 Qy 1539 CCCAAACACCCA 1550
 Db 718 CAAAAACACCA 729

RESULT 8
 BX401619/c

LOCUS	1201 bp	mRNA	linear	EST 13-MAY-2003
DEFINITION	EX401619 Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED			
ACCESSION	EX401619			
KEYWORDS	EX401619.1 GI:30631999			
SOURCE	EST.			
ORGANISM	Homo sapiens (human)			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
AUTHORS	Li, W.B., Gruber, C., Jessee, J. and Polayes, D.			
TITLE	Full-length cDNA libraries and normalization			
JOURNAL	Unpublished (2001)			
COMMENT	Contact: Genoscope Genoscope - Centre National de Sequencage BP 191 91006 EVRY cedex - France Email: seqref@genoscope.cns.fr, web : www.genoscope.cns.fr was normalized. Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 8942.f For more information about this cluster, see http://www.genoscope.cns.fr/ cgi-bin/cluster.cgi?seq=CS0DL005CF10NP1&cluster=8942.f. Contact : Feng Liang Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com/ Invitrogen Corporation 1600 Paraday Avenue Genoscope sequence ID : CS0DL005CF10NP1.			
FEATURES	Location/Qualifiers			
source	1. .1201 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="CS0DL005YL19" /cell_type="B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED" /cell_line="RAMOS CELL LINE" /clone_lib="Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED" /note="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."			
ORIGIN	Query Match 39.3%; Score 670.4; DB 13; Length 1201; Best Local Similarity 84.6%; Pred. No. 2e-161; Matches 798; Conservative 37; Mismatches 92; Indels 16; Gaps 9;			
Qy	747 TAGTCTCTTTCCAGCTTCATCCAGCGCAGTCCGCGCCCTCCAGCAGCGTACTGGGG 806			
Db	958 TRWGTGTTTTCCAGGTTTCATCCAGSAGTCCCGCCCTCCAGCAGGCKTACTGGGG 899			
Qy	807 CTGCTCCCTATAGTCTCTTTATGCGAGCTCCCGAGCAGGAGATGGTCCAGGTGTT-TATC 865			
Db	898 -TGSTCCCTATAGTCTCTTTAT--GGTTCGAGCAGGAGTGGTGGCAGTCTTTATCCC 843			
Qy	866 CCGCCCGCAGGAGTGGGCGCCATCATCGGCAAGAGGGCGCAGCATCAACAGCTCTCC 925			
Db	842 CCGCCCGCAGGATGGGCGCCCTTCWCGGCAAGAGGGCGCAGCACTCAACAGTCTCCC 783			
Qy	926 CGGTTTGCCAGCGGCTCCATCAAGATTGCACACCCGAAACACCTGACTCCAAAGTTCGT 985			
Db	782 CGKTTGCCAGCGYSCYCTGCAAGWTGGACCCCGAAACACTGGATSCAAAKTCTTT 723			
Qy	986 ATGGTTATCATCTGGAACCCAGAGCCCAATTCAGGCTCAGGGAAGATCTATGGC 1045			
Db	722 TTGTTGNTCTCATGGGACCCGACAGGCCCCAATACAGGTTTCAGGAGCAATTTTGTG 663			
Qy	1046 AAATCAAGAGGAG-AACTTCTTTTGGT--CCCAAGGAGGAAGTGAAGCTGGAGACCCAC 1102			
Db	662 AAATCAAGRRRRDAACTCTTTTGGTCCCRARGGGAGTGAAGCTGGAGACCCAC 603			
Qy	1103 ATACGTGTG----CCAGCATCAGCAGCTGGCGGG--TCATTGGCAAGGTGGAAAAACGG 1157			
Db	602 GATTCCGTTGTTCCAGGCACTAGCAGCTGCCGCGCTTANNKTSAAANGGKGAAANCGG 543			

/clone_lib="NIH_MGC_70"

/note="Organ: pancreas; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.1 kb. Library constructed by Life
Technologies."

ORIGIN

```

Query Match      39.2%; Score 669; DB 13; Length 712;
Best Local Similarity 99.3%; Pred. No. 3.4e-161;
Matches 669; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1031 GGAAGATCTATGCGAACTCAAGGAGGAGAACTCTTTGGTCCCAAGGAGGAAGTCAAG 1090
Db 1 GGAAGATCTATGCGAACTCAAGGAGGAGAACTCTTTGGTCCCAAGGAGGAAGTCAAG 60

QY 1091 CTGAGAGCCCATACCTGTGCGAGCATCAGCAGCTGGCCGGGTCAATGGCAAGGTGGA 1150
Db 61 CTGAGAGCCCATACCTGTGCGAGCATCAGCAGCTGGCCGGGTCAATGGCAAGGTGGA 120

QY 1151 AAAACGGTGAACGAGTTGCAGAAATTTGACGGCAGCTGAGGTGGTAGTACCAGAGACCAAG 1210
Db 121 AAAACGGTGAACGAGTTGCAGAAATTTGACGGCAGCTGAGGTGGTAGTACCAGAGACCAAG 180

QY 1211 ACCCTGATGAGAACGACAGGTGATCGTGAATAATCATCGGACATTTCTATGCGAGTCAG 1270
Db 181 ACCCTGATGAGAACGACAGGTGATCGTGAATAATCATCGGACATTTCTATGCGAGTCAG 240

QY 1271 ATGGCTCAACGGGAAGATCCGAGACATCTCGGCCAGGTTAAGCAGCAGCATCAGAAAGGA 1330
Db 241 ATGGCTCAACGGGAAGATCCGAGACATCTCGGCCAGGTTAAGCAGCAGCATCAGAAAGGA 300

QY 1331 CAGAGTAACGAGCCGACGAGCAACGAGGAGTAGTACCGAGGAGTACCGAGGAGTACGAGGGA 1390
Db 301 CAGAGTAACGAGCCGACGAGCAACGAGGAGTAGTACCGAGGAGTACCGAGGAGTACGAGGGA 360

QY 1391 AGGACAAACGAGCGAGAAATCAGAGTGTGCTCTCCCGGAGGAGTACCGAGGAGTACGAGGGA 1450
Db 361 AGGACAAACGAGCGAGAAATCAGAGTGTGCTCTCCCGGAGGAGTACCGAGGAGTACGAGGGA 420

QY 1451 GAATCCGGGACACNTGGCGGGCTGTAGATCAGGTTTGCCACATTTGATGAGAAAGATG 1510
Db 421 GAATCCGGGACACCTGGCGGGCTGTAGATCAGGTTTGCCACATTTGATGAGAAAGATG 480

QY 1511 TTCCAGTGAGAACCTGTATCTTCAGCCCCCAACACCCCAATGGCCCAACACTGT 1570
Db 481 TTCCAGTGAGAACCTGTATCTTCAGCCCCCAACACCCCAATGGCCCAACACTGT 540

QY 1571 NTGCCCTCGGGGTGTGAGAAATTTAGCGCAAGGCACITTTAAACGTGGATTGTTTAA 1630
Db 541 NTGCCCTCGGGGTGTGAGAAATTTAGCGCAAGGCACITTTAAACGTGGATTGTTTAA 600

QY 1631 GAAGCTCTCAGGCCCCCAGAGAGGTTGATCAGCTCAGTGGGAGAAATAAAT 1690
Db 601 GAAGCTCTCAGGCCCCCAGAGAGGTTGATCAGCTCAGTGGGAGAAATAAAT 660

QY 1691 TTCCTTCAGGTTT 1704
Db 661 TTCCTTCAGGTTT 674

```

RESULT 10

BF984962

LOCUS

602308484F1 NIH_MGC_88 Homo sapiens CDNA clone IMAGE:4399556 5',

mRNA sequence.

ACCESSION

BF984962

VERSION

BF984962.1

KEYWORDS

EST.

SOURCE

Homo sapiens

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: ATCC

CDNA Library Preparation: Life Technologies, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LLAM10103 row: h column: 21

High quality sequence stop: 652.

Location/Qualifiers

source

1.896

/organism="Homo sapiens"

/db_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:4399556"

/tissue_type="duodenal adenocarcinoma, cell line"

/lab_host="DH10B (phage-resistant)"

/clone_lib="NIH_MGC_88"

/note="Organ: small intestine; Vector: pCMV-SPORT6;

Site 1: NotI; Site 2: SalI; Cloned unidirectionally;

oligo-dT primed. Average insert size 1.767 kb. Library

enriched for full-length clones and constructed by Life

Technologies. Note: this is a NIH_MGC Library."

ORIGIN

```

Query Match      38.8%; Score 662; DB 10; Length 896;
Best Local Similarity 93.4%; Pred. No. 2.5e-159;
Matches 727; Conservative 0; Mismatches 40; Indels 11; Gaps 3;

QY 583 AGACCTTACCTTTTACAACTGAGAGGACATCATCTGTGAAGGGGCCATCGAGAAATTG 642
Db 1 AGACCTTACCTTTTACAACTGAGAGGACATCATCTGTGAAGGGGCCATCGAGAAATTG 60

QY 643 TTGAGGGCCGAGCAGGAGAAATTAATGAAGAAAGTTGGGAGGCTATGAGAAATCATGTGGC 702
Db 61 TTGAGGGCCGAGCAGGAGAAATTAATGAAGAAAGTTGGGAGGCTATGAGAAATCATGTGGC 120

QY 703 TGCCATGAGC-----TCTCACCTGATCCCTGGCTGAACTGGCTGCTGTAGTCTTTT 756
Db 121 TGCCATGAGCCTGAGTCTCACCTGATCCCTGGCTGAACTGGCTGCTGTAGTCTTTT 180

QY 757 CCCAGCTTCATCCAGCGCAGTCCCGCGCTCCCGAGGAGTACTGGGGCTCTCCCTA 816
Db 181 CCCAGCTTCATCCAGCGCAGTCCCGCGCTCCCGAGGAGTACTGGGGCTCTCTCCCTA 240

QY 817 TAGCTCTTTATGACAGCTCCCGAGCAGGAGATGGTGCAGGTGTTTATCCCCCCCAGGC 876
Db 241 TAGCTCTTTATGACAGCTCCCGAGCAGGAGATGGTGCAGGTGTTTATCCCCCCCAGGC 300

QY 877 AGTGGGCGCCATCATCGGCAAGAGGGGAGCATCAACAGCTCTCCGGTTGCCAG 936
Db 301 AGTGGGCGCCATCATCGGCAAGAGGGGAGCATCAACAGCTCTCCGGTTGCCAG 360

QY 937 CGCTCCATCAAGATTGCACACCCGAAACACCTGACTCCAAAGTTCTGATGGTTATCAT 996
Db 361 CGCTCCATCAAGATTGCACACCCGAAACACCTGACTCCAAAGTTCTGATGGTTATCAT 420

QY 997 CACTGGACCCGAGAGGCCCAATTCAGGCTCAGGGAAGAACTATGGCAAACTCAAGGA 1056
Db 421 CACTGGACCCGAGAGGCCCAATTCAGGCTCAGGGAAGAACTATGGCAAACTCAAGGA 480

QY 1057 GGAGAACTTTTGGTCCCAAGGAGGAAGTGAAGCTGGAGAGCCACATACGTGTGCCAG 1116
Db 481 GGAGAACTTTTGGTCCCAAGGAGGAAGTGAAGCTGGAGAGCCACATACGTGTGCCAG 540

QY 1117 ATCAGAGCTGGCCGGGTGATTTGGCAAGGTGGAAGAAACGGTGAACAGTTCGAGAAATT 1176
Db 541 ATCAGAGCTGGCCGGGTGATTTGGCAAGGTGGAAGAAACGGTGAACAGTTCGAGAAATT 600

```



```
QY 1177 GACGGCAGCTGAGTGGTAGTACCAAGAGACAGACCCCTGATGAGAACGACGAGTCAAT 1236
Db 601 GACGGCAGCTGAGTGGTAGTACCAAGAGACAGACCCCTGATGAGAACGACGAGTCAAT 660
QY 1237 CGTGAAA----AFCATCGGACATTTCTATGCGCAG-TCAGATGGCTCAACGGAAGATCCGA 1291
Db 661 CGTGAAAACATCCATCCGGAACATTTCTATGCGCAGCTCAGATGGCTCAGCGAAGATCCGAG 720
QY 1292 GACATCCTGCCCGCAGCTTAAGCAGCAGCATCAGAGGGGACAGTAACCCAGGCCCGCAGG 1349
Db 721 AACTTCTGCCCCAGGTTAAGCCGCGCAGCATCGAAAGGAGCGGTTCCCGCGCGCGCGG 778

RESULT 11
BX372273/c
LOCUS BX372273 Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED
DEFINITION Homo sapiens cDNA clone CS0DL005YL19 3-PRIME, mRNA sequence.
ACCESSION BX372273
VERSION BX372273.1 GI:30434053
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Eukaryota; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
JOURNAL Full-length cDNA libraries and normalization
COMMENT Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 8942.f For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0BAI043ZG11_CS04119_1&cluster=8942.f.
Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0BAI043ZG11_CS04119_1.
FEATURES
source Location/Qualifiers
1..891
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DL005YL19"
/cell_type="B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED"
/cell_line="RAMOS CELL LINE"
/clone_lib="Homo sapiens B CELLS (RAMOS CELL LINE) COT
25-NORMALIZED"
/notes="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN
Query Match 38.7%; Score 660.2; DB 13; Length 891;
Best Local Similarity 94.1%; Pred. No. 7.1e-159;
Matches 765; Conservative 0; Mismatches 33; Indels 15; Gaps 7;
QY 126 AGAATGGCGCCGAGGGGGGTTTGGCTCTCGGGGTC--AGCCCCGCGCAGGGCTCAGCTGT 183
Db 810 AAAAGTGGCCACAGGGTCTCTGGGCTCTCGGGTCCAGCCCGCCAGGGGATCACTGT 751
QY 184 GGCAGCGGGGGCCCCCAGCCAGC--AGCAGCAAGTGGACATCCCCCTTCGGCTCCT--GGTG 241
Db 750 GCAACGGGGGGGGCCCCCAGCCAAAGCAGCAAGTGGACATCCCCCTTCGGCTCCTCGGGTG 691
QY 242 CCCACCCAGTATGGGTGTCATTATTTGGCAAGAGGGGGCCACCATCCGCAAC-ATCAC 300
Db 690 CCCACAGAAATGTGGGTACCAITATGCGCAAGAGGGGGGGCCCCCATCCCAAAATCAC 631
```

```
QY 301 AAAACAGACCAGTCC-AAGATAGACGTGATAGGAAGAGAAACGAGGTGCAGCTGAAA 359
Db 630 AAAACAGACCAGTCCAAAGATAGACGTGATAGGAAGAGAAACGAGGTGCAGCTGAAA 571
QY 360 AAGCCATCAGTGTGCATCTCCACCCCTGAGGGCTCTCTCCGCTTGTAAAGATGATCTTGG 419
Db 570 AAGCCATCAGTGTGCATCTCCACCCCTGAGGGCTCTCTCCGCTTGTAAAGATGATCTTGG 511
QY 420 AGATTATGCTATAAAGAGCTTAAGGACACCAAAACGGCTGACGAGGTTCCCTCGAAGATCC 479
Db 510 AGATTATGCTATAAAGAGCTTAAGGACACCAAAACGGCTGACGAGTTCCTCTGAGATCC 451
QY 480 TGGCCCAATAAATCTTTAGGGCTCTCATTTGCGAAGAGGACGGAACCTGAAGAAGG 539
Db 450 TGGCCCAATAAATCTTTAGGGCTCTCATTTGCGAAGAGGACGGAACCTGAAGAAGG 391
QY 540 TAGAGCAGATACCGAGACAAAATCCATCTCTCGTTGCAAGACCTTACCTTTTACA 599
Db 390 TAGAGCAGATACCGAGACAAAATCCATCTCTCGTTGCAAGACCTTACCTTTTACA 331
QY 600 ACCCTGAGAGGACCATCACTGTGAAGGGGGCCATCGAGAAATTGTCAGGGCCGAGCAGG 659
Db 330 ACCCTGAGAGGACCATCACTGTGAAGGGGGCCATCGAGAAATTGTCAGGGCCGAGCAGG 271
QY 660 AAATATGAAGAAAGTTCCGAGGCGCTATGAGATGATGTGGCTGCCATGAGC-----T 713
Db 270 AAATATGAAGAAAGTTCCGAGGCGCTATGAGATGATGTGGCTGCCATGAGCTGCAGT 211
QY 714 CTCACCTGATCCCTGGGCTGAACCTGGCTGTGTAGTCTTTTCCCAAGCTTCATCCAGCG 773
Db 210 CTCACCTGATCCCTGGGCTGAACCTGGCTGTGTAGTCTTTTCCCAAGCTTCATCCAGCG 151
QY 774 CAGTCCCGCCGCTCCAGCAGCGTTACTGGGGCTGCTCCCTATAGTCTCTTTATGAGG 833
Db 150 CAGTCCCGCCGCTCCAGCAGCGTTACTGGGGCTGCTCCCTATAGTCTCTTTAT--GG 94
QY 834 CTCGCGACGAGAGATGTGTGAGTGTATTATCCCGCCGCGCAGTGGCGGCCATCATCG 893
Db 93 CTCGCGACGAGAGATGTGTGAGTGTATTATCCCGCCGCGCAGTGGCGGCCATCATCG 34
QY 894 GCAAGAAAGGGGCGAGCAGATCAACAGCTCTCCC 926
Db 33 GCAAGAAAGGGGCGAGCAGATCAACAGCTCTCCC 1

RESULT 12
BX372273
LOCUS BX372273 Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED
DEFINITION Homo sapiens cDNA clone CS0DL005YL19 3-PRIME, mRNA sequence.
ACCESSION BX372273
VERSION BX372273.1 GI:21775837
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Eukaryota; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE NIH-MGC http://mgc.nci.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: CGAP (Stanford)
cDNA Library Preparation: Rubin Laboratory
cDNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM2444 row: a column: 11
High quality sequence stop: 588.
```

FEATURES		Location/Qualifiers	
source		1..955	
		/organism="Homo sapiens"	
		/mol_type="mRNA"	
		/db_xref="taxon:9606"	
		/clones="IMAGE:6269842"	
		/tissue_type="hepatocellular carcinoma, cell line"	
		/lab_host="DH10B (phage-resistant)"	
		/clone_lib="NIH_MGC_100"	
		/note="Organ: liver; Vector: pOTB7; Site: 1: XhoI; Site: 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."	
ORIGIN			
Query Match		38.3%;	Score 653.4; DB 13; Length 955;
Best Local Similarity		98.4%;	Pred. No. 4.2e-157;
Matches	660;	Conservative	0; Mismatches 11; Indels 0; Gaps 0;
QY	30	CCCGGGAGGCATCATGAAGCTGAATGGCCACCAAGTTGGAGAACCATGGCCCTGAAGTCT	89
Db	5	CCAGGCAAGGCATCATGAAGCTGAATGGCCACCAAGTTGGAGAACCATGGCCCTGAAGTCT	64
QY	90	CCTACATCCCGATGACGATAGCAGGAGCTGAGATGGGCGCCGAGGGGGCTTG	149
Db	65	CCTACATCCCGATGACGATAGCAGGAGCTGAGATGGGCGCCGAGGGGGCTTG	124
QY	150	GCTCTCGGGTACGCCCGCCAGCGGTCACTCTGTGACGGGGGCCAGCAAGCAGC	209
Db	125	GCTCTCGGGTACGCCCGCCAGCGGTCACTCTGTGACGGGGGCCAGCAAGCAGC	184
QY	210	AGCAAGTGGACATCCCTTCGGCTCTGGTGGCCACCAAGTATGGGTGCATATTG	269
Db	185	AGCAAGTGGACATCCCTTCGGCTCTGGTGGCCACCAAGTATGGGTGCATATTG	244
QY	270	GCAAGAGGGGGCCACCATCCGCAACATCACAAACAGACCCAGTCCAAAGATAGAGTGC	329
Db	245	GCAAGAGGGGGCCACCATCCGCAACATCACAAACAGACCCAGTCCAAAGATAGAGTGC	304
QY	330	ATAGGAGGAGAACGCGAGGTGAGCTGAAAGCCATCAGTGTGCACTCCACCCCTGAGG	389
Db	305	ATAGGAGGAGAACGCGAGGTGAGCTGAAAGCCATCAGTGTGCACTCCACCCCTGAGG	364
QY	390	GCTGCTCTCCGCTTGTAGATGATCTTGAGATTATGCATAAGAGGCTAAGGACACCA	449
Db	365	GCTGCTCTCCGCTTGTAGATGATCTTGAGATTATGCATAAGAGGCTAAGGACACCA	424
QY	450	AAACGGGTGACGAGTTCCCTTGAAGTCTCTGGCCCAATAAATCTTTAGGGGCTTCA	509
Db	425	AAACGGGTGACGAGTTCCCTTGAAGTCTCTGGCCCAATAAATCTTTAGGGGCTTCA	484
QY	510	TTGGCAAGGAGAACGGAACCTTGAAGAGGTAGAGCAAGTACCGAGACAAATACCA	569
Db	485	TTGGCAAGGAGAACGGAACCTTGAAGAGGTAGAGCAAGTACCGAGACAAATACCA	544
QY	570	TTCTCTGTTGCAAGACCTTACCCCTTACACCCCTGAGAGGACCATCACTGTGAAGGGG	629
Db	545	TTCTCTGTTGCAAGACCTTACCCCTTACACCCCTGAGAGGACCATCACTGTGAAGGGG	604
QY	630	CCATCCGAAATTTCTGAGGGCCGACAGGAATTAATGAAGAGTTCGGAGGCGCTATG	689
Db	605	CCATCCGAAATTTCTGAGGGCCGACAGGAATTAATGAAGAGTTCGGAGGCGCTATG	664
QY	690	AGAAATGATGTG	700
Db	665	CGAATGACGTG	675

RESULT 13
BM006191

LOCUS
DEFINITION

ACCESSION
VERSION

KEYWORDS
SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BM006191 603614013F1 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:5433698 5', mRNA sequence.

BM006191 603614013F1 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:5433698 5', mRNA sequence.

EST. BM006191.1 GI:16520545

Homo sapiens (human)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

NIH-MGC http://mgs.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: CGAP (Stanford)

cDNA Library Preparation: Ling Hong/Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLCMI909 row: j column: 03

High quality sequence stop: 756.

Location/Qualifiers

1..764

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clones="IMAGE:5433698"

/tissue_type="hepatocellular carcinoma, cell line"

/lab_host="DH10B (phage-resistant)"

/clone_lib="NIH_MGC_100"

/note="Organ: liver; Vector: pOTB7; Site: 1: XhoI; Site: 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."

Query Match 37.9%; Score 647.2; DB 12; Length 764;

Best Local Similarity 96.7%; Pred. No. 1.5e-155;

Matches 727; Conservative 0; Mismatches 13; Indels 12; Gaps 6;

QY 30 CCCGGGAGGCATCATGAAGCTGAATGGCCACCAAGTTGGAGAACCATGGCCCTGAAGTCT

Db 7 CCAGGCAAGGCATCATGAAGCTGAATGGCCACCAAGTTGGAGAACCATGGCCCTGAAGTCT

QY 90 CCTACATCCCGATGACGATAGCAGGAGCTGAGATGGGCGCCGAGGGGGCTTG

Db 67 CCTACATCCCGATGACGATAGCAGGAGCTGAGATGGGCGCCGAGGGGGCTTG

QY 150 GCTCTCGGGTACGCCCGCCAGCGGTCACTGTGGCAGCGGGGGGCCAGCAAGCAGC

Db 127 GCTCTCGGGTACGCCCGCCAGCGGTCACTGTGGCAGCGGGGGGCCAGCAAGCAGC

QY 210 AGCAAGTGGACATCCCTTCGGCTCTGGTGGCCACCAAGTATGGGTGCATATTG

Db 186 AGCAAGTGGACATCCCTTCGGCTCTGGTGGCCACCAAGTATGGGTGCATATTG

QY 270 GCAAGAGGGGGCCACCATCCGCAACATCACAAACAGACCCAGTCCAAAGATAGAGTGC

Db 246 GCAAGAGGGGGCCACCATCCGCAACATCACAAACAGACCCAGTCCAAAGATAGAGTGC

QY 330 ATAGGAGGAGAACGCGAGGTGAGCTGAAAGCCATCAGTGTGCACTCCACCCCTGAGG

Db 306 ATAGGAAGGAGAACGCGAGGTGCAGCTGAATAAACCCATCAGTGTGCACCTCCACCCCTTGAG 365
 Qy 389 GGCTGCTCCTCCCTCTGTAA-GATGATCTTGGAGATTATGCATAAAGAGGCTAAGACAC 447
 Db 366 GGCTGCTCCTCCCTCTGTAAAGATGATCTTGGAGATTATGCATAAAGAGGCTAAGACAC 425
 Qy 448 CAAAACGGCTGACGAGGTCCCTCTGAAGATCTTGGGCCCATATAAATCTTTGTAGGGGCTT 507
 Db 426 CAAAACGGCTGACGAGGTCCCTCTGAAGATCTTGGGCCCATATAAATCTTTGTAGGGGCTT 485
 Qy 508 CATTGGCAAGGAGGACGGAACCTGAAGAGGTAGAGGAGATACCGAGACAAAAATCAC 567
 Db 486 CATTGGCAAGGAGGACGGAACCTGAAGAGGTAGAGGAGATACCGAGACAAAAATCAC 545
 Qy 568 CATCTCCTCTGTCGAAGACCTTACCCTTTACAAACCTGAGAGGACCATCACCTGTGAAGG 627
 Db 546 CATCTCCTCTGTCGAAGACCTTACCCTTTACAAACCTGAGAGGACCATCACCTGTGAAGG 605
 Qy 628 GGCCATCGAAGATGTGCGAGGCCGAGCAGGAAATTAATGAAGAAAGTTGCGAGGCTTA 687
 Db 606 GGCCATCGAAGATGTGCGAGGCCGAGCAGGAAATTAATGAAGAAAGTTGCGAGGCTTA 665
 Qy 688 TGAGAATGATGTGGCTGCCATGAGCTCTC-----ACCTGATCCCT-GGCCCTGAACCTG 739
 Db 666 TGAGAATGATGTGGCTGCCATGAGCTCTC-----ACCTGATCCCT-GGCCCTGAACCTG 725
 Qy 740 GCTGCTGTA-GGTCTTTTCCAGCTTCATCCA 770
 Db 726 GCTGCTGTAAGGTCTTTTCCAGCTTCATCCA 757

RESULT 14

B0643920
 LOCUS AGENCOURT_8303966 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6269413
 DEFINITION 5', mRNA sequence.
 ACCESSION B0643920
 VERSION B0643920.1 GI:21768092
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 983)
 AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: CGAP (Stanford)
 cDNA Library Preparation: Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LLCM2442 row: o column: 14
 High quality sequence stop: 491.

FEATURES

source

1..983
 Location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6269413"
 /tissue_type="hepatocellular carcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH_MGC_100"
 /note="Organ: liver; Vector: pOTB7; Site:1: XhoI; Site:2:
 EcoRI; cDNA made by oligo-dT priming. Directionally cloned
 into EcoRI/XhoI sites using the following 5' adaptor:
 GGACACGAG(G). Size-selected >500bp for average insert size
 1.8kb. Library constructed by Ling Hong in the laboratory
 of Gerald M. Rubin (University of California, Berkeley)

using ZAP-cDNA synthesis kit (Stratagene) and Superscript
 II RT (Life Technologies). Note: this is a NIH_MGC
 Library."

ORIGIN

Query Match 36.7%; Score 627.6; DB 13; Length 983;
 Best Local Similarity 96.3%; Pred. No. 1.9e-150;
 Matches 653; Conservative 0; Mismatches 24; Indels 1; Gaps 1;
 Qy 30 CCCGGGAGCCCATCATGAAGCTGAATGGCCACCACTTGGAGAACCATGCCCTGAAGGTCT 89
 Db 5 CCAGGCAAGCCATCATGAAGCTGAATGGCCACCACTTGGAGAACCATGCCCTGAAGGTCT 64
 Qy 90 CTTACATCCCAGATGAGCAGATAGACAGGACCTGAGATGGCGCGCGAGGGGCTTTG 149
 Db 65 CTTACATCCCAGATGAGCAGATAGACAGGACCTGAGATGGCGCGCGAGGGGCTTTG 124
 Qy 150 GCTCTCGGGGTACGCCGCCAGGGCTCACCTGTGGCAGCGGGGCCCCAGCCCAAGCAGC 209
 Db 125 GCTCTCGGGGTACGCCGCCAGGGCTCACCTGTGGCAGCGGGGCCCCAGCCCAAGCAGC 184
 Qy 210 AGCAAGTGGACATCCCCCTTCGGCTCTGGTGCCTGCCACCCAGTATGTGGTGCCATTATTG 269
 Db 185 AGCAAGTGGACATCCCCCTTCGGCTCTGGTGCCTGCCACCCAGTATGTGGTGCCATTATTG 244
 Qy 270 GCAAGGAGGGGGCCACCATCCGCAACATCACAACACAGACCCAGTCCCAAGATGACGTGC 329
 Db 245 GCAAGGAGGGGGCCACCATCCGCAACATCACAACACAGACCCAGTCCCAAGATGACGTGC 304
 Qy 330 ATAGGAAGGAGAACGCGAGGTGCAGCTGAAAAAGCCATCAGTGTGCACCTCCACCCCTGAGG 389
 Db 305 ATAGGAAGGAGAACGCGAGGTGCAGCTGAAAAAGCCATCAGTGTGCACCTCCACCCCTGAGG 364
 Qy 390 GCTGCTCCTCGCTTGTAAAGATGATCTTGGAGATTATGCATAAAGAGGCTAAGACACCA 449
 Db 365 GCTGCTCCTCGCTTGTAAAGATGATCTTGGAGATTATGCATAAAGAGGCTAAGACACCA 424
 Qy 450 AAAAGGCTGACGAGGTTCCTCCTGAAGATCTCTGGGCCCATATAAATCTTGTAGGGGCTTCA 509
 Db 425 AAAAGGCTGACGAGGTTCCTCCTGAAGATCTCTGGGCCCATATAAATCTTGTAGGGGCTTCA 484
 Qy 510 TTGCAAGGAAGGACGGAACCTGAAGAGGTAGAGCAAGATACCGAGACAAAAATCACCA 569
 Db 485 TTGCAAGGAAGGACGGAACCTGAAGAGGTAGAGCAAGATACCGAGACAAAAATCACCA 544
 Qy 570 TCTCCTCCTTCCAGACCTTACCTTTACACCTGAGAGGACCATCACTCTGAAGGGGG 629
 Db 545 TCTCCTCCTTCCAGACCTTACCTTTACACCTGAGAGGACCATCACTCTGAAGGGGG 604
 Qy 630 CC-ATCGAGAAATTGTCAGGGCCGAGCAGGAAATAATGAAGAAAGTTCCGGAGGCCCTAT 688
 Db 605 CCAATCGAGAAATTGTCAGGGCCGAGCAGGAAATAATGAAGAAAGTTCCGGAGGCCCT 664
 Qy 689 GAGAAATGATGGCTGCC 706
 Db 665 AATGAAATGATGGGGGC 682

RESULT 15
 AK088465
 LOCUS AK088465 4100 bp mRNA linear HTC 20-SEP-2003
 DEFINITION Mus musculus 2 days neonate thymus thymic cells cDNA, RIKEN
 full-length enriched library, clone:E430018F06 product:insulin-like
 growth factor 2, binding protein 3, full insert sequence.
 ACCESSION AK088465
 VERSION AK088465.1 GI:26353479
 KEYWORDS HTC; CAP trapper.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1
 AUTHORS Carninci, P. and Hayashizaki, Y.

QY 577 GTTGCAAGACCTTACCTTTACACCTGAGAGACCATCACCTGTGAAGGGGCCATCGA 636
Db |||||
1184 ATTGCAGAACTGACGCTGTACATCCGAAACGACCATTTACAGTGAAGGCAGTGTGA 1243
QY 637 GAATGTTGTGAGGCCCGAGCAGGAAATPAATGAAGAAAGTTCCGGAGGCCCTATCAGAAATGA 696
Db |||||
1244 GACGTGTGCCAAGCGGAGGAGGAAATAATGAAGAGATCAGGGAGTCTTATGAAATGA 1303
QY 697 TGTGGCTGCCATGAGCTCTC-----ACCTGATCCCTGGCCTGAACCTGGCTCTGTAGG 750
Db |||||
1304 TATTGCTTCCATGAATCTTCAAGCAATTTAAATCCCTGGATTAAATCTGAATGCCCTTGGG 1363
QY 751 TCTTTTCCAGCTTCATCCAGCGCAGTCCCGCCGCTCCCGAGCAGGCTTACTGGGCTGC 810
Db |||||
1364 TCTGTTCCCAACCCAGGTCAGGATGCCACCTCCACCTCAGGGCCCCCTTCAACCCCTGAC 1423
QY 811 TCCCTATAGTCTCTTTATGAGGCTCCGAGCAGGAGATGCTGAGGTGTTTATCCCGC 870
Db |||||
1424 TCTTCCCTACCCCAATTTGAGCAATCAGAGACGGAGACTGTGCATCTGTTTATTCGCCG 1483
QY 871 CCAGGCAGTGGGGCCCATCATCGGCAAGAGGGGCGAGCACATCAACAGCTCTCCCGGT 930
Db |||||
1484 CCTGTCGTTGGGCCCATCATTTGGCAGCAGGGCCACACATCAACAGCTTCTCGCTT 1543
QY 931 TGCAGCGCTCCCATCAAGATTGCAOCCACCGAAACACCTGACTCCAAAAGTTCGTATGCT 990
Db |||||
1544 TGCAGGAGCTTCGATTAAAGATCGCTCCAGCCGAAGCACCAGATGCTAAAGTGGGATGCT 1603
QY 991 TATCATCAGTGGACCGCCAGAGGCCCAATTCAGGCTCAGGGAGGATCTATGGCAACT 1050
Db |||||
1604 GATTATCAGTGGACCAACAGAGGCTCAGTTCAAGGCTCAGGGAAGAAATTTATGGAATAAT 1663
QY 1051 CAAGGAGGAGAACTTCTTTGTTCCCAAGGAGGAGTGAAGCTGGAGACCCACATACGTGT 1110
Db |||||
1664 TAAAGAGAAACTTTTGTAGTCTTAAAGAGAGTGAATTTGAAGCTCACATCAGAGT 1723
QY 1111 GCCAGCATCAGCAGCTGGCCGGGTCAATTGCAAGAGGTGGAACACGGTGAACGAGTTGCA 1170
Db |||||
1724 GCGTCTCTTTGCTGCTGGCAGATTATTGGAAAGGAGGCAACCGTGAATGAGCTCCA 1783
QY 1171 GAATTTGACGGCAGCTGAGGTGTAGTACCAGAGAGACCAACCCCTGATGAGAACGCCA 1230
Db |||||
1784 GAGTTTATCAAGTGTGAAGTTGTCTCCCGCTGACCAGACACTGTGAGNATGATCA 1843
QY 1231 GGTCACTCGTGAAATATCATCGGACATTTCTATGCCAGTCAAGTGGCTCAACGGGAAGATCCG 1290
Db |||||
1844 AGTAGTTGTCAAAATAACTGCCCACTTCTATGCTTCCAGGTTGCCCGAGGAAAAATTC 1903
QY 1291 AGACATCCTGCCCGGTTAAGCAGCAGCATCAGAG 1327
Db |||||
1904 GGAATTTCTGACTCAGGTAAGCAGCACCCAGCAGCAG 1940

Search completed: July 13, 2004, 21:40:34
Job time : 4603 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.
OM nucleic - nucleic search, using sw model
Run on: July 14, 2004, 01:45:09 ; Search time 1662 Seconds
(without alignments)
5008.780 Million cell updates/sec

Title: US-09-270-437D-5
Perfect score: 1708
Sequence: 1 agggagcgtgcgcacgccc.....atttccttcagggttttaaaa 1708

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 3183909 seqs, 2436941669 residues

Word size : 0

Total number of hits satisfying chosen parameters: 6367818

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications NA:
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:
13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq2:
14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:
15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:
16: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:
17: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:
18: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:
19: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1708	100.0	1708	9 US-09-899-651-5	Sequence 5, Appli
2	1672	97.9	1946	9 US-09-899-651-7	Sequence 7, Appli
3	741	43.4	1182	9 US-09-764-864-327	Sequence 327, App
4	588	34.4	1129	9 US-09-764-864-745	Sequence 745, App
5	365	21.4	1738	16 US-10-062-674-2082	Sequence 2082, Ap
6	125	7.3	831	13 US-10-027-632-157364	Sequence 157364,
7	125	7.3	831	16 US-10-027-632-157364	Sequence 157364,
8	101	5.9	762	16 US-10-027-632-20264	Sequence 20264, A
9	60	3.5	60	10 US-09-908-975-15619	Sequence 15619, A
10	60	3.5	60	10 US-09-908-975-31745	Sequence 31745, A
11	60	3.5	60	10 US-09-908-975-31873	Sequence 31873, A
12	50	2.9	50	16 US-10-131-827-2939	Sequence 2939, Ap
13	31	1.8	2224	9 US-09-873-637-1	Sequence 1, Appli
14	29	1.7	412	16 US-10-062-674-1244	Sequence 1244, Ap

Sequence 38, Appl
Sequence 36, Appl
Sequence 30184, A
Sequence 16926, A
Sequence 16926, A
Sequence 9333, Ap
Sequence 15683, A
Sequence 141668, A
Sequence 97191, A
Sequence 11930, A
Sequence 10752, A
Sequence 47612, A
Sequence 46916, A
Sequence 16656, A
Sequence 270463, A
Sequence 270463, A
Sequence 2356, Ap
Sequence 115479, A
Sequence 115479, A
Sequence 70, Appl
Sequence 70, Appl
Sequence 6811, Ap
Sequence 347, App
Sequence 347, App
Sequence 347, App
Sequence 347, App
Sequence 347, App
Sequence 483, App
Sequence 347, App
Sequence 478, App
Sequence 478, App

ALIGNMENTS

RESULT 1

US-09-899-651-5
; Sequence 5, Application US/09899651
; Patent No. US20020111470A1
; GENERAL INFORMATION:
; APPLICANT: Chen, Yao-Tseng
; APPLICANT: Tsang, Solam
; APPLICANT: Stockert, Elisabeth
; APPLICANT: Jager, Elke
; APPLICANT: Knuth, Alexander
; APPLICANT: Old, Lloyd J.
; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Encoding Cancer
; TITLE OF INVENTION: Associated
; TITLE OF INVENTION: Antigen, The Antigens Per Se, And Uses Thereof
; FILE REFERENCE: LUD 5538
; CURRENT APPLICATION NUMBER: US/09/899,651
; CURRENT FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: US/09/061,709
; PRIOR FILING DATE: 1998-04-17
; NUMBER OF SEQ ID NOS: 8
; SEQ ID NO 5
; LENGTH: 1708
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
US-09-899-651-5

Query Match 100.0%; Score 1708; DB 9; Length 1708;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1708; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGAGCGTGCACCGCCCGCCAGTTACCCGGGAGCCATCATGAGCTGAATGGCCA 60
Db 1 AGGAGCGTGCACCGCCCGCCAGTTACCCGGGAGCCATCATGAGCTGAATGGCCA 60
QY 61 CCAGTTGGAGAACCATGTCCTTGAAGTCTCTTACATCCCGGATGACGATAGCAGGG 120

Db 61 CCAAGTTGGAGAACCATGCTGAGAGTCTCTATACATCCCGATGAGCAGATAGCACAGG 120
Qy 121 ACCTGAGAAATGGGGCGCGAGGGGCTTTGGCTCTCGGGGTGAGCCCGCCAGGGCTCAC 180
Db 121 ACCTGAGAAATGGGGCGCGAGGGGCTTTGGCTCTCGGGGTGAGCCCGCCAGGGCTCAC 180
Qy 181 TGTGGCAGCGGGGCGCCAGCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 240
Db 181 TGTGGCAGCGGGGCGCCAGCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 240
Qy 241 GCCACCCAGTATGTGGGTGCGCATTTATGGCAAGAGGGGGGCGACACATCGCGAATCAC 300
Db 241 GCCACCCAGTATGTGGGTGCGCATTTATGGCAAGAGGGGGGCGACACATCGCGAATCAC 300
Qy 301 AAAACAGACCCAGTCCCAAGATAGACGTGCATAGGAAGAGAAACGCGAGGTGACGTGAAA 360
Db 301 AAAACAGACCCAGTCCCAAGATAGACGTGCATAGGAAGAGAAACGCGAGGTGACGTGAAA 360
Qy 361 AGCCATCAGTGTGCACTCCACCCCTGAGGGCTGCTCTCCCGCTTGTAAAGATGATCTTGA 420
Db 361 AGCCATCAGTGTGCACTCCACCCCTGAGGGCTGCTCTCCCGCTTGTAAAGATGATCTTGA 420
Qy 421 GATTATGCATAAAGAGCTAAGGACACCAAAACGGCTGACGAGGTTCCTCTGAAGATCCT 480
Db 421 GATTATGCATAAAGAGCTAAGGACACCAAAACGGCTGACGAGGTTCCTCTGAAGATCCT 480
Qy 481 GGGCCATAAATCTTTGTAGGGGCTCTCATTTGGCAAGAGAGCGAAACCTGGAAGAGT 540
Db 481 GGGCCATAAATCTTTGTAGGGGCTCTCATTTGGCAAGAGAGCGAAACCTGGAAGAGT 540
Qy 541 AGAGCAAGATACCGAGACAAAATACCATCTCTGTGTCGAGAGACCTTACCCCTTACAA 600
Db 541 AGAGCAAGATACCGAGACAAAATACCATCTCTGTGTCGAGAGACCTTACCCCTTACAA 600
Qy 601 CCCTGAGAGGACCATCACTGTGAAGGGGCGCATCGAGAAATTTGTGAGGGCGCGAGGAG 660
Db 601 CCCTGAGAGGACCATCACTGTGAAGGGGCGCATCGAGAAATTTGTGAGGGCGCGAGGAG 660
Qy 661 AATAATGAAGAGTTCGGAGGCGCTATGAGAAATGATGTGGCTGCGCATGAGCTCTCACCT 720
Db 661 AATAATGAAGAGTTCGGAGGCGCTATGAGAAATGATGTGGCTGCGCATGAGCTCTCACCT 720
Qy 721 GATCCCTGGCTGAACCTGCTGCTGTAGTCTTTCCAGCTTTCATCCAGCGCAGTCCC 780
Db 721 GATCCCTGGCTGAACCTGCTGCTGTAGTCTTTCCAGCTTTCATCCAGCGCAGTCCC 780
Qy 781 GCGGCTCCCGAGCAGGCTTACTGGGCTGCTCCCTATAGTCTCTTTATCGAGCTCCCGA 840
Db 781 GCGGCTCCCGAGCAGGCTTACTGGGCTGCTCCCTATAGTCTCTTTATCGAGCTCCCGA 840
Qy 841 GCAGGAGATGGTCAGTGTATATCCCGCCAGGAGTGGGCGCCATCATGGCGAAGAA 900
Db 841 GCAGGAGATGGTCAGTGTATATCCCGCCAGGAGTGGGCGCCATCATGGCGAAGAA 900
Qy 901 GGGGCGAGCACAATCAAGCAGCTCTCCCGGTTTGGCAGCGCTCTCATCAAGATTGCAACCC 960
Db 901 GGGGCGAGCACAATCAAGCAGCTCTCCCGGTTTGGCAGCGCTCTCATCAAGATTGCAACCC 960
Qy 961 CGAAACACCTGACTCCAAAGTTCGTATGTTATCATCTGAGCCCGCAGAGGCCCAATT 1020
Db 961 CGAAACACCTGACTCCAAAGTTCGTATGTTATCATCTGAGCCCGCAGAGGCCCAATT 1020
Qy 1021 CAAGGCTCAGGGAAGATCTATGGCAAACTCAAGGAGGAGAACTCTTTGGTCCCAAGGA 1080
Db 1021 CAAGGCTCAGGGAAGATCTATGGCAAACTCAAGGAGGAGAACTCTTTGGTCCCAAGGA 1080
Qy 1081 GGAAGTGAAGCTGGAGACCCACATACGTGTGCGAGCATCAGAGCTGGCCGGTCAATTGG 1140
Db 1081 GGAAGTGAAGCTGGAGACCCACATACGTGTGCGAGCATCAGAGCTGGCCGGTCAATTGG 1140
Qy 1141 CAAAGCTGGAAGAAACGGTGAACAGTTGAGAAATTTGAGGAGCTGAGGTGCTAGTACC 1200

Db 1141 CAAAGGTGGAAGAAACGGTGAACAGTTGACAGAAATTTGACGGCAGCTGAGGTGGTAGTACC 1200
Qy 1201 AAGAGACGAGACCCCTGATGAGAACGACGACGATCATCGTGAAGATCATCGACATTTCTA 1260
Db 1201 AAGAGACGAGACCCCTGATGAGAACGACGACGATCATCGTGAAGATCATCGACATTTCTA 1260
Qy 1261 TGCCAGTCAGATGGCTCAACGGAAGATCCGAGACATCTCGCCGCCAGGTTAAGCAGCAGCA 1320
Db 1261 TGCCAGTCAGATGGCTCAACGGAAGATCCGAGACATCTCGCCGCCAGGTTAAGCAGCAGCA 1320
Qy 1321 TCAGAAGGAGCAGAGTAAACAGGCCCGAGGCAACGAGGAAGTGAACAGCCCTCCCTGTCC 1380
Db 1321 TCAGAAGGAGCAGAGTAAACAGGCCCGAGGCAACGAGGAAGTGAACAGCCCTCCCTGTCC 1380
Qy 1381 CTTNGAGTCCAGGCAACCAACGGGCGAGAAATCGAGAGTGTCTCTCCCGCGCAGGCCCTGA 1440
Db 1381 CTTNGAGTCCAGGCAACCAACGGGCGAGAAATCGAGAGTGTCTCTCCCGCGCAGGCCCTGA 1440
Qy 1441 GAATGAGTGGAAATCCGGGACACNTGGCCCGGCTGTAGATCAGGTTTGCCCACTTGATT 1500
Db 1441 GAATGAGTGGAAATCCGGGACACNTGGCCCGGCTGTAGATCAGGTTTGCCCACTTGATT 1500
Qy 1501 GAGAAGATGTTCCAGTGAGGAACCTGTCTNTCAGCCCCCAACACCCACCCCAATTGGC 1560
Db 1501 GAGAAGATGTTCCAGTGAGGAACCTGTCTNTCAGCCCCCAACACCCACCCCAATTGGC 1560
Qy 1561 CCAACACTGTTGCCCCCTCGGGGTGTGAGAAATNTAGCGCAAGGCACTTTTAAACGTGG 1620
Db 1561 CCAACACTGTTGCCCCCTCGGGGTGTGAGAAATNTAGCGCAAGGCACTTTTAAACGTGG 1620
Qy 1621 ATTGTTTAAAGAGTCTCCAGGCCCAACGAGAGGTGGATCACCTCAGTGGGGAAGA 1680
Db 1621 ATTGTTTAAAGAGTCTCCAGGCCCAACGAGAGGTGGATCACCTCAGTGGGGAAGA 1680
Qy 1681 AAAATAAAATTTCTTCAGGTTTAAAAA 1708
Db 1681 AAAATAAAATTTCTTCAGGTTTAAAAA 1708

RESULT 2

US-09-899-651-7

; Sequence 7, Application US/09899651

; Patent No. US20020111470A1

; GENERAL INFORMATION:

; APPLICANT: Chen, Rao-Tseng

; APPLICANT: Gure, Ali

; APPLICANT: Tsang, Solam

; APPLICANT: Stockert, Elisabeth

; APPLICANT: Jager, Elke

; APPLICANT: Knuth, Alexander

; APPLICANT: Old, Lloyd J.

; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Encoding Cancer

; TITLE OF INVENTION: Associated

; TITLE OF INVENTION: Antigen, The Antigens Per Se, And Uses Thereof

; FILE REFERENCE: LUD 5538

; CURRENT APPLICATION NUMBER: US/09/899,651

; CURRENT FILING DATE: 2001-07-06

; PRIOR APPLICATION NUMBER: US/09/061,709

; PRIOR FILING DATE: 1998-04-17

; NUMBER OF SEQ ID NOS: 8

; SEQ ID NO 7

; LENGTH: 1946

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

US-09-899-651-7

Query Match 97.9%; Score 1672; DB 9; Length 1946;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1672; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 37 AGCCATCATGAAGTGAATCGCCACGAGTTGGAGAACCATGCGCTGAAGTCTCTACAT 96


```
Db 275 AGCATCATGAAGCTGAATGGCCACACAGTTGGAGAACCATGCCCTGAAGGTCTCTACAT 334
Qy 97 CCCGATGACAGATAGCA CAGGGA CTTGAGATGGCGCGCGAGGGGCTTTGGCTCTCG 156
Db 335 CCCGATGACAGATAGCA CAGGGA CTTGAGATGGCGCGCGAGGGGCTTTGGCTCTCG 394
Qy 157 GGTTCAGCCCGCCAGGGCTCACTGTGGCAGCGGGGCCCCAGCCAGCAGCAAGT 216
Db 395 GGTTCAGCCCGCCAGGGCTCACTGTGGCAGCGGGGCCCCAGCCAGCAGCAAGT 454
Qy 217 GGCATCCCCCTTTGGCTCTCGTGGCCACCCAGTATGTGGTGCCATTATTTGGCAAGGA 276
Db 455 GGCATCCCCCTTTGGCTCTCGTGGCCACCCAGTATGTGGTGCCATTATTTGGCAAGGA 514
Qy 277 GGGGCCACCATCGGAACATCACAAACAGACCCAGTCCAGATAGAGTGCATAGGAA 336
Db 515 GGGGCCACCATCGGAACATCACAAACAGACCCAGTCCAGATAGAGTGCATAGGAA 574
Qy 337 GAGAAACGAGGTGCAGCTGAAAGACCATCAGTGTGCACTCCACCCCTGAGGGCTGCTC 396
Db 575 GAGAAACGAGGTGCAGCTGAAAGACCATCAGTGTGCACTCCACCCCTGAGGGCTGCTC 634
Qy 397 CTCGGCTTGAAGTATCTTGAGATTAATGATTAAGAGGCTTAAGGACACCAAAACGGC 456
Db 635 CTCGGCTTGAAGTATCTTGAGATTAATGATTAAGAGGCTTAAGGACACCAAAACGGC 694
Qy 457 TGACGAGTTCCCTGAAGATCTGGGCCATAATACTTTGTAGGGGCTCTCATTTGGCAA 516
Db 695 TGACGAGTTCCCTGAAGATCTGGGCCATAATACTTTGTAGGGGCTCTCATTTGGCAA 754
Qy 517 GGAAGACGGAACCTGGAAGAGGTAGAGCAAGATACCGAGACAAATAACCATCTCTCTC 576
Db 755 GGAAGACGGAACCTGGAAGAGGTAGAGCAAGATACCGAGACAAATAACCATCTCTCTC 814
Qy 577 GTTGCAAGACTTACCTTTACACCTGAGAGGACCATCACTGTGAAGGGGCCATCGA 636
Db 815 GTTGCAAGACTTACCTTTACACCTGAGAGGACCATCACTGTGAAGGGGCCATCGA 874
Qy 637 GAATTGTTGAGGGCCGAGCAGGAATAATGAAGAAAGTTCCGGAGGCCCTATGAGAAATGA 696
Db 875 GAATTGTTGAGGGCCGAGCAGGAATAATGAAGAAAGTTCCGGAGGCCCTATGAGAAATGA 934
Qy 697 TGTGGCTGCATGAGCTCTACCTGATCCCTGGCCTGAACCTGGCTGCTGTAGGTCTTTT 756
Db 935 TGTGGCTGCATGAGCTCTACCTGATCCCTGGCCTGAACCTGGCTGCTGTAGGTCTTTT 994
Qy 757 CCCAGCTTCATCCAGCGAGTCCCGCCGCTCCAGCAGCGGTACTGGGCTGCTCCCTA 816
Db 995 CCCAGCTTCATCCAGCGAGTCCCGCCGCTCCAGCAGCGGTACTGGGCTGCTCCCTA 1054
Qy 817 TAGTCTCTTTATGAGGCTCCGAGCAGGAGTGTGAGGTGTATATCCCGCCAGGC 876
Db 1055 TAGTCTCTTTATGAGGCTCCGAGCAGGAGTGTGAGGTGTATATCCCGCCAGGC 1114
Qy 877 AGTGGCGCCATATCCGCAAGAGGGGCGAGCATCAACAGCTCTCCCGGTTTGGCAG 936
Db 1115 AGTGGCGCCATATCCGCAAGAGGGGCGAGCATCAACAGCTCTCCCGGTTTGGCAG 1174
Qy 937 GCGCTCCATCAAGATTGACACCCGAAACACCTGACTCCAAAGTTCGTATGTTATCAT 996
Db 1175 GCGCTCCATCAAGATTGACACCCGAAACACCTGACTCCAAAGTTCGTATGTTATCAT 1234
Qy 997 CACTGGACCGCAGAGGCCCAATTCAGGCTCAGGGAAGATCTATGGCAAACTCAAGGA 1056
Db 1235 CACTGGACCGCAGAGGCCCAATTCAGGCTCAGGGAAGATCTATGGCAAACTCAAGGA 1294
Qy 1057 GGAGAACTTCTTTGGTCCCAAGGAGGAAGTGAAGCTGGAGACCCACATACGTGTGCCAGC 1116
Db 1295 GGAGAACTTCTTTGGTCCCAAGGAGGAAGTGAAGCTGGAGACCCACATACGTGTGCCAGC 1354
Qy 1117 ATCAGCAGCTGGCGGGTCAATTGGCAAAAGGTGGAAAAACGGTGAACCGAGTTGAGAAATTT 1176
Db 1355 ATCAGCAGCTGGCGGGTCAATTGGCAAAAGGTGGAAAAACGGTGAACCGAGTTGAGAAATTT 1414
```

```
Qy 1177 GACGCGACCTGAGTGTAGTAGTACCAAGAGACCAACCCCTGATGAGAACACCAAGTTCAT 1236
Db 1415 GACGCGACCTGAGTGTAGTAGTACCAAGAGACCAACCCCTGATGAGAACACCAAGTTCAT 1474
Qy 1237 GGTGAAATCATCGACATTTCTATGCCAGTCAAGTGGCTCAACGGGAAGATCCGAGACAT 1296
Db 1475 GGTGAAATCATCGACATTTCTATGCCAGTCAAGTGGCTCAACGGGAAGATCCGAGACAT 1534
Qy 1297 CTTGCCCGCAGGTTAAGCAGAGCATCAGAGGGAAGATTAACAGGCCCGAGGACCGAG 1356
Db 1535 CTTGCCCGCAGGTTAAGCAGAGCATCAGAGGGAAGATTAACAGGCCCGAGGACCGAG 1594
Qy 1357 GAAGTGACAGCCCTCTCTGTCCTTNGAGTCCAGGACAAACACGGGACAGAAATCCGAGA 1416
Db 1595 GNAGTGACAGCCCTCTCTGTCCTTNGAGTCCAGGACAAACACGGGACAGAAATCCGAGA 1654
Qy 1417 GTGTGCTCTCCCGCAGCGCTGAGATGAGTGGGAATCCGGGACACATGTCGGGCGGTG 1476
Db 1655 GTGTGCTCTCCCGCAGCGCTGAGATGAGTGGGAATCCGGGACACATGTCGGGCGGTG 1714
Qy 1477 TAGATCAGGTTTGGCCACTTGATTGAGAAAGATGTTCCAGTGAGGAACCTGTATCTNCA 1536
Db 1715 TAGATCAGGTTTGGCCACTTGATTGAGAAAGATGTTCCAGTGAGGAACCTGTATCTNCA 1774
Qy 1537 GCCCAAAACACCCACCCCAATTTGGCCCAACACTGNTGCCCTCGGGGTGTGAGAAATNT 1596
Db 1775 GCCCAAAACACCCACCCCAATTTGGCCCAACACTGNTGCCCTCGGGGTGTGAGAAATNT 1834
Qy 1597 AGCGAAGGCACTTTTAAACGTGGATGTTTAAAGAGCTCTCCAGGCCCCACCAAGAGG 1656
Db 1835 AGCGAAGGCACTTTTAAACGTGGATGTTTAAAGAGCTCTCCAGGCCCCACCAAGAGG 1894
Qy 1657 GTGGATCACACCTCAGTGGGAAGAAATAAAATTTCCCTCAGGTTTTTAAA 1708
Db 1895 GTGGATCACACCTCAGTGGGAAGAAATAAAATTTCCCTCAGGTTTTTAAA 1946
```

RESULT 3

```
US-09-764-864-327
; Sequence 327, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 327
; LENGTH: 1182
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-864-327
```

Query Match 43.4%; Score 741; DB 9; Length 1182;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 991; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```
Qy 713 TCTACCTGATCCCTGGCTGAACCTGCTGCTGTAGGTCTTTTCCAGCTTCATCCAGC 772
Db 136 TCTACCTGATCCCTGGCTGAACCTGCTGCTGTAGGTCTTTTCCAGCTTCATCCAGC 195
Qy 773 GCAGTCCCGCCCTCCAGCAGCGTTACTGGGGTGTCTCCCTATAGTCTCTTTATGAG 832
Db 196 GCAGTCCCGCCCTCCAGCAGCGTTACTGGGGTGTCTCCCTATAGTCTCTTTATGAG 255
Qy 833 GCTCCCGCAGCAGAGTGTGAGGTGTTTATCCCGCCCGCAGCTGGCGCCATCATC 892
Db 256 GCTCCCGCAGCAGAGTGTGAGGTGTTTATCCCGCCCGCAGCTGGCGCCATCATC 315
```

```
QY 893 GGCAGAGAGGGGAGGACATCAACAGCTCTCCCGGTTTGCAGCGCTCCATCAAGATT 952
DB 316 GGCAGAGAGGGGAGGACATCAACAGCTCTCCCGGTTTGCAGCGCTCCATCAAGATT 375
QY 953 GCACACCCGAAACACCTGACTCAAAAGTTGATGTTTATATCACTGAGCGCCAGAG 1012
DB 376 GCACACCCGAAACACCTGACTCAAAAGTTGATGTTTATCATCACTGAGCGCCAGAG 435
QY 1013 GCCCAATTCAAGGCTCAGGAGAGATCTATGCGAACTCAAGAGGAGAACTTCTTTGGT 1072
DB 436 GCCCAATTCAAGGCTCAGGAGAGATCTATGCGAACTCAAGAGGAGAACTTCTTTGGT 495
QY 1073 CCCAAGGAGAGAGTGAAGCTGGAGACCCACATACGTGTGCCAGCATCAGCAGCTGGCCGG 1132
DB 496 CCCAAGGAGAGAGTGAAGCTGGAGACCCACATACGTGTGCCAGCATCAGCAGCTGGCCGG 555
QY 1133 GTCAATTGGCAAGGTGAAAGAACCGGTGAACGAGTTGAGAAATTTGACGGAGCTGAGGTG 1192
DB 556 GTCAATTGGCAAGGTGAAAGAACCGGTGAACGAGTTGAGAAATTTGACGGAGCTGAGGTG 615
QY 1193 GTAGTACCAAGAGACAGACCCCTGATGAGAACGACAGCTCATCTGAAATCATCGGA 1252
DB 616 GTAGTACCAAGAGACAGACCCCTGATGAGAACGACAGCTCATCTGAAATCATCGGA 675
QY 1253 CATTTCTATGCCAGTCAAGTGGTCAACGGAAGATCCGAGACATCTGCGCCAGGTTAAG 1312
DB 676 CATTTCTATGCCAGTCAAGTGGTCAACGGAAGATCCGAGACATCTGCGCCAGGTTAAG 735
QY 1313 CAGCAGCATCAGAGGACAGAGTAAACAGGCCCGCAGCGAGGAGTGAACAGCCCT 1372
DB 736 CAGCAGCATCAGAGGACAGAGTAAACAGGCCCGCAGCGAGGAGTGAACAGCCCT 795
QY 1373 CCTGTCTCTTNGAGTCCAGGACAAACGCGGAGAGAAATCGAGAGTGTCTCTCCCGGC 1432
DB 796 CCTGTCTCTTNGAGTCCAGGACAAACGCGGAGAGAAATCGAGAGTGTCTCTCCCGGC 855
QY 1433 AGGCTGAGATGATGGGATCGGGACACNTGGCGCGGCTGTAGATCAGTTTGGCC 1492
DB 856 AGGCTGAGATGATGGGATCGGGACACNTGGCGCGGCTGTAGATCAGTTTGGCC 915
QY 1493 ACTTGATTGAGAAAGATGTTCCAGTGAAGAACCTGATCTNTCAGCCCCAAACACCCACC 1552
DB 916 ACTTGATTGAGAAAGATGTTCCAGTGAAGAACCTGATCTNTCAGCCCCAAACACCCACC 975
QY 1553 CAATTGGCCCAACACTGNTGCCCTCGCGGCTGTAGAAATTTAGCGCAAGCACTTTT 1612
DB 976 CAATTGGCCCAACACTGNTGCCCTCGCGGCTGTAGAAATTTAGCGCAAGCACTTTT 1035
QY 1613 AAACGTGATGTTTAAAGAGCTCTCCAGGCCCCACAGAGGGTGGATCACACCTCAG 1672
DB 1036 AAACGTGATGTTTAAAGAGCTCTCCAGGCCCCACAGAGGGTGGATCACACCTCAG 1095
QY 1673 TGGGAGAAAAATAAAATTTCTTCAGGTTTAAAA 1708
DB 1096 TGGGAGAAAAATAAAATTTCTTCAGGTTTAAAA 1131
```

RESULT 4

```
US-09-764-864-745
; Sequence 745, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PZ23
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 745
; LENGTH: 1129
; TYPE: DNA
```

i: ORGANISM: Homo sapiens
US-09-764-864-745

Query Match 34.4%; Score 588; DB 9; Length 1129;
Best Local Similarity 99.2%; Pred. No. 3.9e-297;
Matches 988; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

```
QY 713 TCTCACTGATCCCTGGCGCTGAACCTGGCTGCTAGTCTCTTTCCAGCTTCATCCAGC 772
DB 103 TCTCACTGATCCCTGGCGCTGAACCTGGCTGCTAGTCTCTTTCCAGCTTCATCCAGC 162
QY 773 GCAGTCCCGCGCGCTCCCGAGCAGCGTTACTGGGGCTGCTCCCTATAGTCTCTTTATGCG 832
DB 163 GCAGTCCCGCGCGCTCCCGAGCAGCGTTACTGGGGCTGCTCCCTATAGTCTCTTTATGCG 222
QY 833 GCTCCCGAGCAGGAGATGGTGCAGGTGTTTATCCCGCCAGGAGTGGGGCGCATCATC 892
DB 223 GCTCCCGAGCAGGAGATGGTGCAGGTGTTTATCCCGCCAGGAGTGGGGCGCATCATC 282
QY 893 GGCAGAGAGGGGCGAGCATCAACAGCTCTCCCGGTTTGCAGCGCTCCATCAAGATT 952
DB 283 GGCAGAGAGGGGCGAGCATCAACAGCTCTCCCGGTTTGCAGCGCTCCATCAAGATT 342
QY 953 GCACACCCGAAACACCTGACTCCAAAGTTCGTATGGKTATCATYACTGGACCGCCARAG 1012
DB 343 GCACACCCGAAACACCTGACTCCAAAGTTCGTATGGKTATCATYACTGGACCGCCARAG 402
QY 1013 GCCCAATTCAGGCTCAGGGAGAGATCTATGGCAAACTCAAGGAGGAGAACTTCTTTGGT 1072
DB 403 GCCCAATTCAGGCTCAGGGAGAGATCTATGGCAAACTCAAGGAGGAGAACTTCTTTGGT 462
QY 1073 CCCAAGGAGAGTGAAGCTGGAGACCCACATACGTGTGCCAGCATCAGCAGCTGGCCGG 1132
DB 463 CCCAAGGAGAGTGAAGCTGGAGACCCACATACGTGTGCCAGCATCAGCAGCTGGCCGG 522
QY 1133 GTCAATGGCAAGGTGGAAGAACCGTGAACGAGTTGCAGAAATTTGACGGCAGCTGAGGTG 1192
DB 523 GTCAATGGCAAGGTGGAAGAACCGTGAACGAGTTGCAGAAATTTGACGGCAGCTGAGGTG 582
QY 1193 GTAGTACCAAGAGACAGACCCCTGATGAGAACGACAGCTCATCTGTAATAATCATCGGA 1252
DB 583 GTAGTACCAAGAGACAGACCCCTGATGAGAACGACAGCTCATCTGTAATAATCATCGGA 642
QY 1253 CATTTCTATGCGAGTCAAGTGGCTCAACGGAAGATCCGAGACATCTTGGCCCGAGTTAAG 1312
DB 643 CATTTCTATGCGAGTCAAGTGGCTCAACGGAAGATCCGAGACATCTTGGCCCGAGTTAAG 702
QY 1313 CAGCAGCATCAGAGGAGCAGAGTAAACAGGCCCGCAGGAGGAGTGAACAGCCCT 1372
DB 703 CAGCAGCATCAGAGGAGCAGAGTAAACAGGCCCGCAGGAGGAGTGAACAGCCCT 762
QY 1373 CCCTGTCCCTTNGAGTCCAGGACAAACGCGGAGAGAAATCGAGAGTGTCTCTCCCGGC 1432
DB 763 CCCTGTCCCTTNGAGTCCAGGACAAACGCGGAGAGAAATCGAGAGTGTCTCTCCCGGC 822
QY 1433 AGGCTGAGATGATGGGATCCGGACACNTGGCGCGGCTGTAGATCAGGTTGGCC 1492
DB 823 AGGCTGAGATGATGGGATCCGGACACCTGGCGCGGCTGTAGATCAGGTTGGCC 882
QY 1493 ACTTGATTGAGAAAGATGTTCCAGTGAAGAACCTGATCTNTCAGCCCCAAACACCCACC 1552
DB 883 ACTTGATTGAGAAAGATGTTCCAGTGAAGAACCTGATCTNTCAGCCCCAAACACCCACC 942
QY 1553 CAATTGGCCCAACACTGNTGCCCTCGCGGCTGTAGAAATTTAGCGCAAGCACTTTT 1612
DB 943 CAATTGGCCCAACACTGNTGCCCTCGCGGCTGTAGAAATTTAGCGCAAGCACTTTT 1002
QY 1613 AAACGTGATGTTTAAAGAGAGTCTCCAGGCCCCACAGAGGGTGGATCACACCTCAG 1672
DB 1003 AAACGTGATGTTTAAAGAGAGTCTCCAGGCCCCACAGAGGGTGGATCACACCTCAG 1062
QY 1673 TGGGAGAAAAATAAAATTTCTTCAGGTTTAAAA 1708
```

```
Db 1063 TGGGAGAAAAATAAAATTCCTTCAGGTTTAA 1098
RESULT 5
US-10-062-674-2082/c
; Sequence 2082, Application US/10062674
; Publication No. US20040005559A1
; GENERAL INFORMATION:
; APPLICANT: Loring, Jeanne F.; Kaser, Matthew R.
; TITLE OF INVENTION: MARKERS OF NEURONAL DIFFERENTIATION AND MORPHOGENESIS
; FILE REFERENCE: FA-0026-1 CIP
; CURRENT APPLICATION NUMBER: US/10/062,674
; CURRENT FILING DATE: 2002-01-30
; PRIOR APPLICATION NUMBER: US 09/625,102
; PRIOR FILING DATE: 2000-07-24
; NUMBER OF SEQ ID NOS: 2217
; SOFTWARE: PERL Program
; SEQ ID NO 2082
; LENGTH: 1738
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20040005559A1 457674.1
US-10-062-674-2082
Query Match 21.4%; Score 365; DB 16; Length 1738;
Best Local Similarity 100.0%; Pred. No. 2.8e-180; Indels 0; Gaps 0;
Matches 365; Conservative 0; Mismatches 0;
Qy 37 AGCCATCATGAAGCTGAATGCGCCACCCAGTTGGAGAACCATGCCCTGAAGGTCCTCTACAT 96
Db 1728 AGCCATCATGAAGCTGAATGCGCCACCCAGTTGGAGAACCATGCCCTGAAGGTCCTCTACAT 1669
Qy 97 CCCCGATGACGAGATAGACAGGACCTGAGATGGCGCCGAGGGGGCTTTGGCTCTCG 156
Db 1668 CCCCGATGACGAGATAGACAGGACCTGAGATGGCGCCGAGGGGGCTTTGGCTCTCG 1609
Qy 157 GGGTCACCCCGCCAGGGCTCACCTGTGGCAGCGGGGGCCCGCCAGCAGCAAGT 216
Db 1608 GGGTCACCCCGCCAGGGCTCACCTGTGGCAGCGGGGGCCCGCCAGCAGCAAGT 1549
Qy 217 GGACATCCCGCTTGGCTCTGTGCCCCACCCAGTATGTGGTGCCATTATTGGCAAGGA 276
Db 1548 GGACATCCCGCTTGGCTCTGTGCCCCACCCAGTATGTGGTGCCATTATTGGCAAGGA 1489
Qy 277 GGGGGCCACCATCGCAACATCACAAACAGACCCAGTCCCAAGATAGACGTGCATAGGAA 336
Db 1488 GGGGGCCACCATCGCAACATCACAAACAGACCCAGTCCCAAGATAGACGTGCATAGGAA 1429
Qy 337 GGAGAACGCGAGGTGACGTGAAAGCCATCAGTGTGCACTCCACCCCTGAGGGCTGCTC 396
Db 1428 GGAGAACGCGAGGTGACGTGAAAGCCATCAGTGTGCACTCCACCCCTGAGGGCTGCTC 1369
Qy 397 CTCGG 401
Db 1368 CTCGG 1364
RESULT 6
US-10-027-632-157364
; Sequence 157364, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 157364
; LENGTH: 831
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-157364
Query Match 7.3%; Score 125; DB 16; Length 831;
Best Local Similarity 100.0%; Pred. No. 1.6e-54; Indels 0; Gaps 0;
Matches 125; Conservative 0; Mismatches 0;
Qy 454 GGCTGACGAGGTTCCCTGGAAGATCCTGGCCCAATAAATTAACCTTTGAGGGCGTCTCATGG 513
Db 344 GGCTGACGAGGTTCCCTGGAAGATCCTGGCCCAATAAATTAACCTTTGAGGGCGTCTCATGG 403
Qy 514 CAAGGAAGGACGGACCTGAAGAGGTAGACGACGACGACGACGACGACGACGACGACGACGAC 573
Db 404 CAAGGAAGGACGGACCTGAAGAGGTAGACGACGACGACGACGACGACGACGACGACGACGAC 463
Qy 574 CTCGT 578
Db 464 CTCGT 468
RESULT 7
US-10-027-632-157364
; Sequence 157364, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 157364
; LENGTH: 831
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-157364
Query Match 7.3%; Score 125; DB 16; Length 831;
Best Local Similarity 100.0%; Pred. No. 1.6e-54; Indels 0; Gaps 0;
Matches 125; Conservative 0; Mismatches 0;
```

QY 454 GGCTGACGAGTTCCTCCCTGAAGATCCTGGCCATAATAACTTTGTAGGGGCTCTCATTTGG 513
|
|
|
Db 344 GGCTGACGAGTTCCTCCCTGAAGATCCTGGCCATAATAACTTTGTAGGGGCTCTCATTTGG 403
|
|
|
QY 514 CAAGGAGGAGCGGAACTTCTGAGAGGTAGAGCAAGATACCGAGACAAAATACCATCTC 573
|
|
|
Db 404 CAAGGAGGAGCGGAACTTCTGAGAGGTAGAGCAAGATACCGAGACAAAATACCATCTC 463
|
|
|
QY 574 CTCGT 578
|
|
|
Db 464 CTCGT 468
|
|
|

RESULT 8
US-10-027-632-20264
; Sequence 20264, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR FILING DATE: 2002-04-30
; PRIOR FILING DATE: 2000-07-12 US 60/218,006
; PRIOR FILING DATE: 2000-07-12 US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR FILING DATE: 2000-03-29 US 60/193,483
; PRIOR FILING DATE: 2000-03-29 US 60/185,218
; PRIOR FILING DATE: 2000-02-24 US 60/167,363
; PRIOR FILING DATE: 1999-11-23 US 60/156,358
; PRIOR FILING DATE: 1999-09-28 US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20264
; LENGTH: 762
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-20264

Query Match 5.9%; Score 101; DB 16; Length 762;
Best Local Similarity 100.0%; Pred. No. 6.2e-42;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 354 CTGAAAAGCCATCAGTGTGCATCCACCCTGAGGGCTCTCTCCGCTTTGTAAGATGA 413
|
|
|
Db 1 CTGAAAAGCCATCAGTGTGCATCCACCCTGAGGGCTCTCTCCGCTTTGTAAGATGA 60
|
|
|
QY 414 TCTTGAGATTATGCATAAGAGCGCTAAGACACCAAACG 454
|
|
|
Db 61 TCTTGAGATTATGCATAAGAGCGCTAAGACACCAAACG 101
|
|
|

RESULT 9
US-09-908-975-15619
; Sequence 15619, Application US/09908975
; Publication No. US20030165843A1
; GENERAL INFORMATION:
; APPLICANT: SHOSHAN, Avi
; APPLICANT: WASSERMAN, Alon
; APPLICANT: MINTZ, Eli
; APPLICANT: MINTZ, Liat
; APPLICANT: FAIGLER, Simchon
; TITLE OF INVENTION: OLIGONUCLEOTIDE LIBRARY FOR DETECTING RNA TRANSCRIPTS AND SPLICE
; FILE REFERENCE: 36688-0005
; CURRENT APPLICATION NUMBER: US/09/908,975

; CURRENT FILING DATE: 2001-07-20
; PRIOR APPLICATION NUMBER: US 60/287,724
; PRIOR FILING DATE: 2001-05-02
; PRIOR APPLICATION NUMBER: US 60/221,607
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 32337
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 15619
; LENGTH: 60
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-908-975-15619

Query Match 3.5%; Score 60; DB 10; Length 60;
Best Local Similarity 100.0%; Pred. No. 1.9e-20;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1243 AATCATCGGACATTTCTATGCCAGTCAGATGGCTCAACGGAAGATCCGAGACATCCTGGC 1302
|
|
|
Db 1 AATCATCGGACATTTCTATGCCAGTCAGATGGCTCAACGGAAGATCCGAGACATCCTGGC 60
|
|
|

RESULT 10
US-09-908-975-31745
; Sequence 31745, Application US/09908975
; Publication No. US20030165843A1
; GENERAL INFORMATION:
; APPLICANT: SHOSHAN, Avi
; APPLICANT: WASSERMAN, Alon
; APPLICANT: MINTZ, Eli
; APPLICANT: MINTZ, Liat
; APPLICANT: FAIGLER, Simchon
; TITLE OF INVENTION: OLIGONUCLEOTIDE LIBRARY FOR DETECTING RNA TRANSCRIPTS AND SPLICE
; FILE REFERENCE: 36688-0005
; CURRENT APPLICATION NUMBER: US/09/908,975
; CURRENT FILING DATE: 2001-07-20
; PRIOR APPLICATION NUMBER: US 60/287,724
; PRIOR FILING DATE: 2001-05-02 US 60/221,607
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 32337
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 31745
; LENGTH: 60
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-908-975-31745

Query Match 3.5%; Score 60; DB 10; Length 60;
Best Local Similarity 100.0%; Pred. No. 1.9e-20;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1470 CGGCTGTAGATCAGGTTTCCCACTTGATGAGAAAGATGTTCCAGTAGGAAACCTGA 1529
|
|
|
Db 1 CGGCTGTAGATCAGGTTTCCCACTTGATGAGAAAGATGTTCCAGTAGGAAACCTGA 60
|
|
|

RESULT 11
US-09-908-975-31873
; Sequence 31873, Application US/09908975
; Publication No. US20030165843A1
; GENERAL INFORMATION:
; APPLICANT: SHOSHAN, Avi
; APPLICANT: WASSERMAN, Alon
; APPLICANT: MINTZ, Eli
; APPLICANT: MINTZ, Liat
; APPLICANT: FAIGLER, Simchon
; TITLE OF INVENTION: OLIGONUCLEOTIDE LIBRARY FOR DETECTING RNA TRANSCRIPTS AND SPLICE
; FILE REFERENCE: 36688-0005
; CURRENT APPLICATION NUMBER: US/09/908,975
; CURRENT FILING DATE: 2001-07-20

```
; PRIOR APPLICATION NUMBER: US 60/287,724
; PRIOR FILING DATE: 2001-05-02
; PRIOR APPLICATION NUMBER: US 60/221,607
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 32337
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 31873
; LENGTH: 60
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-908-975-31873

Query Match
Best Local Similarity 3.5%; Score 60; DB 10; Length 60;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1470 CGGGCTGTAGATCAGGTTTCCCACTTGATTGAGAAAGATGTTCCAGTGAGGAACCTGA 1529
Db 1 CGGGCTGTAGATCAGGTTTCCCACTTGATTGAGAAAGATGTTCCAGTGAGGAACCTGA 60

RESULT 12
US-10-131-827-2939
; Sequence 2939, Application US/10131827
; Publication No. US20040009479A1
; GENERAL INFORMATION:
; APPLICANT: Fy, Kirk
; APPLICANT: Wohlgenuth, Jay
; APPLICANT: Woodward, Robert
; APPLICANT: Ly, Ngoc
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING AND MONITORING AUTOIMMUNE
; TITLE OF INVENTION: CHRONIC INFLAMMATORY DISEASES
; FILE REFERENCE: 506612000120
; CURRENT FILING DATE: 2002-09-06
; PRIOR APPLICATION NUMBER: US/10/131,827
; PRIOR FILING DATE: 2002-09-06
; PRIOR APPLICATION NUMBER: US 10/006,290
; PRIOR FILING DATE: 2001-10-22
; PRIOR APPLICATION NUMBER: US 60/296,764
; PRIOR FILING DATE: 2001-06-08
; NUMBER OF SEQ ID NOS: 9090
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2939
; LENGTH: 50
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-131-827-2939

Query Match
Best Local Similarity 2.9%; Score 50; DB 16; Length 50;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1652 AGAGGGTGGATCACACCTCAGTGGGAAGAAAATAAAATTCCTTCAGGT 1701
Db 1 AGAGGGTGGATCACACCTCAGTGGGAAGAAAATAAAATTCCTTCAGGT 50

RESULT 13
US-09-873-637-1
; Sequence 1, Application US/09873637
; Patent No. US20020061543A1
; GENERAL INFORMATION:
; APPLICANT: Ross, Jeffrey
; TITLE OF INVENTION: THE C-MYC CODING REGION DETERMINANT-BINDING PROTEIN
; TITLE OF INVENTION: (CRD-BP) AND ITS NUCLEIC ACID SEQUENCE
; FILE REFERENCE: 960296.95131
; CURRENT APPLICATION NUMBER: US/09/873,637
; CURRENT FILING DATE: 2001-06-04
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2224
; TYPE: DNA
; ORGANISM: Mus musculus
```

```
US-09-873-637-1

Query Match
Best Local Similarity 1.8%; Score 31; DB 9; Length 2224;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 66 TGGAGAACCATGCGCCTGAAGGTCTCTCATCAT 96
Db 561 TGGAGAACCATGCGCCTGAAGGTCTCTCATCAT 591

RESULT 14
US-10-062-674-1244
; Sequence 1244, Application US/10062674
; Publication No. US20040005559A1
; GENERAL INFORMATION:
; APPLICANT: Loring, Jeanne F.; Kaser, Matthew R.
; TITLE OF INVENTION: MARKERS OF NEURONAL DIFFERENTIATION AND MORPHOGENESIS
; FILE REFERENCE: PA-0036-1 CIP
; CURRENT APPLICATION NUMBER: US/10/062,674
; CURRENT FILING DATE: 2002-01-30
; PRIOR APPLICATION NUMBER: US 09/525,102
; PRIOR FILING DATE: 2000-07-24
; NUMBER OF SEQ ID NOS: 2217
; SOFTWARE: PERL Program
; SEQ ID NO 1244
; LENGTH: 412
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: GenBank ID No. US20040005559A1 g1595304
US-10-062-674-1244

Query Match
Best Local Similarity 1.7%; Score 29; DB 16; Length 412;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 599 AACCTGTAGAGGACCATCATCTGTGAAGGG 627
Db 308 AACCTGTAGAGGACCATCATCTGTGAAGGG 336

RESULT 15
US-09-873-637-38/c
; Sequence 38, Application US/09873637
; Patent No. US20020061543A1
; GENERAL INFORMATION:
; APPLICANT: Ross, Jeffrey
; TITLE OF INVENTION: THE C-MYC CODING REGION DETERMINANT-BINDING PROTEIN
; TITLE OF INVENTION: (CRD-BP) AND ITS NUCLEIC ACID SEQUENCE
; FILE REFERENCE: 960296.95131
; CURRENT APPLICATION NUMBER: US/09/873,637
; CURRENT FILING DATE: 2001-06-04
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 38
; LENGTH: 24
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Oligonucleotide primer
US-09-873-637-38

Query Match
Best Local Similarity 1.4%; Score 24; DB 9; Length 24;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 508 CATTGGCAAGGAGGACCGAACCT 531
Db 24 CATTGGCAAGGAGGACCGAACCT 1
```

Wed Jul 14 08:56:08 2004

us-09-270-437d-5.rnpb

Page 8

Search completed: July 14, 2004, 05:27:47
Job time : 1666 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 13, 2004, 21:40:40 ; Search time 700 Seconds
(without alignments)
10365.603 Million cell updates/sec

Title: US-09-270-437D-5
Perfect score: 1708
Sequence: 1 agggagcgcgcacgcgc.....atttccttcagggttttaaaa 1708

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 3373863 seqs, 2124099041 residues

Word size : 0

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_29Jan04:*
1: Geneseq1980s:*
2: Geneseq1990s:*
3: Geneseq2000s:*
4: Geneseq2001as:*
5: Geneseq2001bs:*
6: Geneseq2002s:*
7: Geneseq2003as:*
8: Geneseq2003bs:*
9: Geneseq2003cs:*
10: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1708	100.0	1708	3	Aaz36151 DNA encod
2	1672	97.9	1946	3	Aaz36153 An altern
3	741	43.4	1182	4	Aas26148 Human CDN
4	741	43.4	1182	7	Abx73489 Human nov
5	741	43.4	2780	4	Aak94782 Human ful
6	588	34.4	1129	4	Aas26566 Human CDN
7	588	34.4	1129	7	Abx73907 Human nov
8	414	24.2	833	4	Aak91969 Human CDN
9	414	24.2	833	4	Aak93655 Human CDN
10	60	3.5	60	6	Abn42871 Human spl
11	60	3.5	60	6	Abn59125 Human spl
12	60	3.5	60	6	Abn58997 Human spl
13	50	2.9	50	6	Aaz03248 Human leu
14	31	1.8	2224	2	Aaz10617 cDNA enco
15	24	1.4	24	2	Aaz10625 PCR prime
16	23	1.3	23	2	Aaz10623 PCR prime
17	23	1.3	23	3	Aaz36158 PCR prime
18	23	1.3	65	6	Abn57436 Mouse spl
19	22	1.3	329	6	Abt10006 Human bre
20	22	1.3	465	1	Aan50077 Soybean h
21	20	1.2	456	5	Aas86149 DNA encod
22	20	1.2	526	3	AAC00956 Human sec
23	20	1.2	899	4	Aah05615 Human CDN

C	24	20	1.2	1465	3	AAF18051	Aaf18051 Lung canc
C	25	20	1.2	1553	9	ADC31862	Adc31862 Human nov
C	26	20	1.2	1553	9	ADe09844	Adc09844 Novel DNA
	27	20	1.2	1740	3	AAC66035	Aac66035 Human lun
	28	20	1.2	1740	6	ABL49254	Abi49254 Human lun
	29	20	1.2	1740	6	ABQ92440	Abq92440 Human lun
	30	20	1.2	1740	8	ADA28437	Ada28437 Human lun
	31	20	1.2	1743	6	ABL49299	Abi49299 Human lun
	32	20	1.2	1743	6	ABL49297	Abi49297 Human lun
	33	20	1.2	1743	6	ABQ92483	Abq92483 Human lun
	34	20	1.2	1743	6	ABQ92485	Abq92485 Human lun
	35	20	1.2	1743	8	ADA28540	Ada28540 Recombina
	36	20	1.2	1743	8	ADA28537	Ada28537 Recombina
	37	20	1.2	1764	6	ABL49283	Abi49283 Human lun
	38	20	1.2	1764	6	ABQ92469	Abq92469 Human lun
	39	20	1.2	1764	8	ADA28518	Ada28518 Recombina
	40	20	1.2	1985	5	AAS76779	Aas76779 DNA encod
	41	20	1.2	2963	4	AAL17630	Aah17630 Human CDN
C	42	20	1.2	3055	6	ABX97087	Abx97087 Human NOV
C	43	20	1.2	4108	5	ABV24751	Abv24751 Human pro
	44	20	1.2	4159	3	Aaz36150	Aaz36150 DNA encod
	45	20	1.2	4181	3	AAC65900	Aac65900 Human lun

ALIGNMENTS

RESULT 1
AAZ36151
ID AAZ36151 standard; DNA; 1708 BP.
XX
AC AAZ36151;
XX
DT 11-FEB-2000 (first entry)
XX
DE DNA encoding cancer associated antigen KOC-2.
XX
KW Cancer associated antigen; KOC-2; cancer; vaccine; CT7; ss.
XX
OS Homo sapiens.
XX
PN WO9954738-A1.
XX
PD 28-OCT-1999.
XX
PF 16-MAR-1999; 99WO-US005766.
XX
PR 17-APR-1998; 98US-00061709.
XX
PA (LUDW-) LUDWIG INST CANCER RES.

Chen Y, Gure A, Tsang S, Stockert E, Jager E, Knuth A, Old LJ;
WPI; 2000-013284/01.
Nucleotides representing cancer-associated genes, used to develop
products for the diagnosis, monitoring and treatment of cancers.

Claim 55; Page 40; 44pp; English.

The present sequence represents a cancer associated antigen gene
designated KOC-2. The specification also describes a cancer associated
antigen designated CT7. The CT7 polynucleotide was isolated from SK-MEL-
37 melanoma cells. The polypeptide has some homology with MAGE-10,
limited to about 210 carboxy terminal amino acids. The amino terminal of
the protein has a repetitive pattern, with repeats rich in serine,
proline, glutamine and leucine, and an almost invariable core of the
peptide given in AAY43877. The CT7 polypeptide can be processed to
peptides which provoke lysis by cytolytic T cells. The polynucleotides
and polypeptides can be used for treating a cancerous condition and
screening for or diagnosing cancerous conditions. The cancer associated
antigens can be used as an immunogenic or vaccine composition with an
adjuvant, e.g. a cytokine, a saponin, or granulocyte macrophage-colony

CC stimulating factor (GM-CSF)
XX Sequence 1708 BP; 447 A; 469 C; 473 G; 314 T; 0 U; 5 Other;
SQ

Query Match 100.0%; Score 1708; DB 3; Length 1708;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1708; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGAGCGCTGCGCACCGCCCTTACCCCGGGAGCCATCATGAGCTGATGCGCA 60
DB 1 AGGAGCGCTGCGCACCGCCCTTACCCCGGGAGCCATCATGAGCTGATGCGCA 60

QY 61 CCAGTTGAGAACCATGCTGAGAGTCTCTACATCCCGAGTACAGATAGCAGGG 120
DB 61 CCAGTTGAGAACCATGCTGAGAGTCTCTACATCCCGAGTACAGATAGCAGGG 120

QY 121 ACCTGAGATGGGCGCGAGGGGCTTGGCTTCGGGGTACGCCCGCAGGGCTCAC 180
DB 121 ACCTGAGAAATGGGCGCGAGGGGCTTGGCTTCGGGGTACGCCCGCAGGGCTCAC 180

QY 181 TGTGGCAGCGGGGCCCCAGCAGCAGCAGTGGACATCCCGCTTCGGCTCTCTGGT 240
DB 181 TGTGGCAGCGGGGCCCCAGCAGCAGCAGTGGACATCCCGCTTCGGCTCTCTGGT 240

QY 241 GCCCAGCAGTATGTGGTGGCTATTATGGCAAGAGGGGGCCACCATCGCAACATC 300
DB 241 GCCCAGCAGTATGTGGTGGCTATTATGGCAAGAGGGGGCCACCATCGCAACATC 300

QY 301 AAAACAGACCCAGTCCAGATAGACGTCATAGAGAGAGAGAGAGAGTGCAGTCA 360
DB 301 AAAACAGACCCAGTCCAGATAGACGTCATAGAGAGAGAGAGAGTGCAGTCA 360

QY 361 AGCCATCAGTGTGCTCCACCCCTGAGGGCTGCTCTCGCTTGTAGATGATCTTGA 420
DB 361 AGCCATCAGTGTGCTCCACCCCTGAGGGCTGCTCTCGCTTGTAGATGATCTTGA 420

QY 421 GATTATGCATAAGAGCTTAAGACACCAAAAGGCTGACGAGTTCCTGAGATCTT 480
DB 421 GATTATGCATAAGAGCTTAAGACACCAAAAGGCTGACGAGTTCCTGAGATCTT 480

QY 481 GGCCATAATACTTTGTAGGGGCTCTCATTTGCAAGAGAGAGAGAGAGTGCAGT 540
DB 481 GGCCATAATACTTTGTAGGGGCTCTCATTTGCAAGAGAGAGAGAGTGCAGT 540

QY 541 AGAGCAAGATACCGAGCAAAATACATCTCTCTGTTGCAAGACCTTACCTTTAC 600
DB 541 AGAGCAAGATACCGAGCAAAATACATCTCTCTGTTGCAAGACCTTACCTTTAC 600

QY 601 CCTGAGAGGAGCCATCACTGTGAAGGGGCGCCATCGAGATTTGTCAGGGCGAGC 660
DB 601 CCTGAGAGGAGCCATCACTGTGAAGGGGCGCCATCGAGATTTGTCAGGGCGAGC 660

QY 661 AATAATGAAGAAAGTTGCGAGGCTTATGAGATGATGTGGTGCATCAGCTCTCAC 720
DB 661 AATAATGAAGAAAGTTGCGAGGCTTATGAGATGATGTGGTGCATCAGCTCTCAC 720

QY 721 GATCCCTGGCTGAACTGCTGCTGAGTCTTTTCCAGCTTCACTCAGCGCTGCC 780
DB 721 GATCCCTGGCTGAACTGCTGCTGAGTCTTTTCCAGCTTCACTCAGCGCTGCC 780

QY 781 GCCGCTCCAGAGGCTTACTGGGCTGCTCCCTATAGCTCTTTATGAGGCTCCCG 840
DB 781 GCCGCTCCAGAGGCTTACTGGGCTGCTCCCTATAGCTCTTTATGAGGCTCCCG 840

QY 841 GCAGGAGTGGTCAGAGTGTATCCCGCCAGGAGTGGGCGCATCATCGGCAAGAA 900
DB 841 GCAGGAGTGGTCAGAGTGTATCCCGCCAGGAGTGGGCGCATCATCGGCAAGAA 900

QY 901 GGCGCAGCAGCATCAACAGCTCTCCCGGTTTGCAGCGCTCCATCAAGATTGCA 960
DB 901 GGCGCAGCAGCATCAACAGCTCTCCCGGTTTGCAGCGCTCCATCAAGATTGCA 960

QY 961 CGNAACACTGACTCCAAAGTTGATGTTATCATCTGAGACCGCAGAGGCCCAATT 1020

DB 961 CGAAACACTGACTCCAAAGTTGATGTTATCATCTGAGCGCCAGAGGCCCAATT 1020
QY 1021 CAAGGCTCAGGGAAGATCTATGGCAAACTCAAGAGGAGAACTCTTTTGGTCCCAAGGA 1080
DB 1021 CAAGGCTCAGGGAAGATCTATGGCAAACTCAAGAGGAGAACTCTTTTGGTCCCAAGGA 1080

QY 1081 GGAAGTGAAGCTGGAGACCCCATACAGTGTGCCAGCATCAGCAGCTGSCCGGGTCATTGG 1140
DB 1081 GGAAGTGAAGCTGGAGACCCCATACAGTGTGCCAGCATCAGCAGCTGSCCGGGTCATTGG 1140

QY 1141 CAAAGGTCGAAAAACGGTGAACAGGTTGCAGAAATTTGACGGCAGCTGAGTGGTAGTACC 1200
DB 1141 CAAAGGTCGAAAAACGGTGAACAGGTTGCAGAAATTTGACGGCAGCTGAGTGGTAGTACC 1200

QY 1201 AAGAGACAGACCCCTGATGAGAACGACAGCTCATCTGTAATAATCATCGGACATTTCTA 1260
DB 1201 AAGAGACAGACCCCTGATGAGAACGACAGCTCATCTGTAATAATCATCGGACATTTCTA 1260

QY 1261 TGCCAGTCAGATGGCTCAACGGAAGATCCGAGACATCTGCGCCAGGTTAAGCAGCAGCA 1320
DB 1261 TGCCAGTCAGATGGCTCAACGGAAGATCCGAGACATCTGCGCCAGGTTAAGCAGCAGCA 1320

QY 1321 TCAGAGGACAGAGTAAACAGGCCCAGGCAACGAGGAAGTGAACAGCCCTCCCTCTGCC 1380
DB 1321 TCAGAGGACAGAGTAAACAGGCCCAGGCAACGAGGAAGTGAACAGCCCTCCCTCTGCC 1380

QY 1381 CTTTGAGTCCAGGCAACACCGGCGAGAAATCGAGAGTGTCTCTCCCGCAGGCTCTGA 1440
DB 1381 CTTTGAGTCCAGGCAACACCGGCGAGAAATCGAGAGTGTCTCTCCCGCAGGCTCTGA 1440

QY 1441 GAATGAGTGGGAATCCCGGACACNTGGGCGGGCTGTAGATCAGGTTTGCCACTTGATT 1500
DB 1441 GAATGAGTGGGAATCCCGGACACNTGGGCGGGCTGTAGATCAGGTTTGCCACTTGATT 1500

QY 1501 GAGAAAGATGTTCCAGTGAAGAACCTTGATCTNTCAGCCCAACACCCACCCCAATTGGC 1560
DB 1501 GAGAAAGATGTTCCAGTGAAGAACCTTGATCTNTCAGCCCAACACCCACCCCAATTGGC 1560

QY 1561 CCAACACTGTTGCGGCTCGGGGTGTGAGAAATTTAGCCAGGCACTTTTAAACGTGG 1620
DB 1561 CCAACACTGTTGCGGCTCGGGGTGTGAGAAATTTAGCCAGGCACTTTTAAACGTGG 1620

QY 1621 ATTGTTAAAGAGCTCTCCAGGCCCCCAAGAGGGTGGATCACCTCAGTGGGAAGA 1680
DB 1621 ATTGTTAAAGAGCTCTCCAGGCCCCCAAGAGGGTGGATCACCTCAGTGGGAAGA 1680

QY 1681 AAAATAAAATTTCCCTCAGGTTTAAAA 1708
DB 1681 AAAATAAAATTTCCCTCAGGTTTAAAA 1708

RESULT 2

AAZ36153
ID AAZ36153 standard; DNA; 1946 BP.
XX AAZ36153;
AC AAZ36153;
XX
DT 11-FEB-2000 (first entry)
XX
DE An alternative form of DNA encoding cancer associated antigen KOC-2.
XX
KW Cancer associated antigen; KOC-2; cancer; vaccine; CT7; ss.
XX
OS Homo sapiens.
XX
FN WO954738-A1.
XX
PD 28-OCT-1999.
XX
PF 16-MAR-1999; 99WO-US005766.
XX
PR 17-APR-1998; 98US-00061709.

XX PA (LUDW-) LUDWIG INST CANCER RES.
XX PI Chen Y, Gure A, Tsang S, Stockert E, Jager E, Knuth A, Old LJ;
XX DR WPI; 2000-013284/01.
XX PT Nucleotides representing cancer-associated genes, used to develop
XX PS products for the diagnosis, monitoring and treatment of cancers.
XX SS Claim 55; Page 42; 44pp; English.
XX CC The present sequence represents an alternative form of a cancer
XX CC associated antigen gene designated KOC-2. The specification also
XX CC describes a cancer associated antigen designated CT7. The CT7
XX CC polynucleotide was isolated from SK-MEL-37 melanoma cells. The
XX CC polypeptide has some homology with MAGP-10, limited to about 210 carboxy
XX CC terminal amino acids. The amino terminal of the protein has a repetitive
XX CC pattern, with repeats rich in serine, proline, glutamine and leucine, and
XX CC an almost invariable core of the peptide given in AAY43877. The CT7
XX CC polypeptide can be processed to peptides which provoke lysis by cytolytic
XX CC T cells. The polynucleotides and polypeptides can be used for treating a
XX CC cancerous condition and screening for or diagnosing cancerous conditions.
XX CC The cancer associated antigens can be used as an immunogenic or vaccine
XX CC composition with an adjuvant, e.g. a cytokine, a saponin, or granulocyte
XX CC macrophage-colony stimulating factor (GM-CSF)
SQ Sequence 1946 BP; 502 A; 528 C; 553 G; 358 T; 0 U; 5 Other;
Query Match 97.9%; Score 1672; DB 3; Length 1946;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1672; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 37 AGCATCATGAAGCTGAATGGCCACCAAGTTGGAGAACCATGCCCTGAAGTCTCTACAT 96
DB 275 AGCATCATGAAGCTGAATGGCCACCAAGTTGGAGAACCATGCCCTGAAGTCTCTACAT 334
QY 97 CCCGATGACAGATACACAGGACCTGAGATGGCGCCGAGGGGCTTGGCTCTCG 156
DB 335 CCCGATGACAGATACACAGGACCTGAGATGGCGCCGAGGGGCTTGGCTCTCG 394
QY 157 GGGTACGCCCCGAGGGCTCACCTGTGGCAGCGGGGGCCCCAGCCAGCAGCAAGT 216
DB 395 GGGTACGCCCCGAGGGCTCACCTGTGGCAGCGGGGGCCCCAGCCAGCAGCAAGT 454
QY 217 GGACATCCCTTCGGCTCTGGTCCACCCAGTATGGGTGCCATTATTGGCAAGGA 276
DB 455 GGACATCCCTTCGGCTCTGGTCCACCCAGTATGGGTGCCATTATTGGCAAGGA 514
QY 277 GGGGGCCACCATCCGCAACATCAAAAACAGACCCAGTCCAAAGATAGACGTGATAGGAA 336
DB 515 GGGGGCCACCATCCGCAACATCAAAAACAGACCCAGTCCAAAGATAGACGTGATAGGAA 574
QY 337 GGAAGACGAGGTGCGAGCTGAAAGACCATCATGTGCACTCCACCCCTGAGGGCTGCTC 396
DB 575 GGAAGACGAGGTGCGAGCTGAAAGACCATCATGTGCACTCCACCCCTGAGGGCTGCTC 634
QY 397 CTCGGCTTGAAGATGATCTTGGAGATTATGCAATAAAGGCTTAAGGACACCAAAACGGC 456
DB 635 CTCGGCTTGAAGATGATCTTGGAGATTATGCAATAAAGGCTTAAGGACACCAAAACGGC 694
QY 457 TGACGAGTTCCCTGGAAGTCTCTGGCCCAATAATACTTTGTAGGGCGTCTCATTTGGCAA 516
DB 695 TGACGAGTTCCCTGGAAGTCTCTGGCCCAATAATACTTTGTAGGGCGTCTCATTTGGCAA 754
QY 517 GGAAGACGGAACCTGGAAGAGGTAGAGCAAGATACCGAGCAAAAATCACCATCTCTCTC 576
DB 755 GGAAGACGGAACCTGGAAGAGGTAGAGCAAGATACCGAGCAAAAATCACCATCTCTCTC 814
QY 577 GTTGCAAGACCTTACCCCTTTACACCCCTGAGAGGACCATCCTGTGAAGGGGGCCATCGA 636
DB 815 GTTGCAAGACCTTACCCCTTTACACCCCTGAGAGGACCATCCTGTGAAGGGGGCCATCGA 874

QY 637 GAATTGTTGACGGCCGAGCAGGAAATATATGAAGAAAGTTTCGGGAGGCTATGAGAATCA 696
DB 875 GAATTGTTGACGGCCGAGCAGGAAATATATGAAGAAAGTTTCGGGAGGCTATGAGAATCA 934
QY 697 TGTGGCTGCATGAGCTCTACCTCATCCCTGGCTGAAACCTGGCTGTGTAGGCTTTT 756
DB 935 TGTGGCTGCATGAGCTCTACCTCATCCCTGGCTGAAACCTGGCTGTGTAGGCTTTT 994
QY 757 CCCAGCTTCATCCAGCGAGTCCCGCCCTCCAGCAGCGTTACTCTGGGGCTGCTCCCTA 816
DB 995 CCCAGCTTCATCCAGCGAGTCCCGCCCTCCAGCAGCGTTACTCTGGGGCTGCTCCCTA 1054
QY 817 TAGCTCCTTTATGACGGCTCCGAGCAGGAGATGTTGAGGTGTTTATCCCGCCGAGGC 876
DB 1055 TAGCTCCTTTATGACGGCTCCGAGCAGGAGATGTTGAGGTGTTTATCCCGCCGAGGC 1114
QY 877 AGTGGGGCCCATCATCGGCAGAGGGGAGGAGCAGACATCAAAAGCTCTCCCGTTTCCAG 936
DB 1115 AGTGGGGCCCATCATCGGCAGAGGGGAGGAGCAGACATCAAAAGCTCTCCCGTTTCCAG 1174
QY 937 CGCTCCCATCAAGATTGCACCAACCAACCTGACTCCAAAGTTCGTATGGTTATCAT 996
DB 1175 CGCTCCCATCAAGATTGCACCAACCAACCTGACTCCAAAGTTCGTATGGTTATCAT 1234
QY 997 GACTGAGCGCCAGAGGGCCCAATTCAAGGCTCAGGGAAGAAATCTATGGAACAACTCAAGGA 1056
DB 1235 GACTGAGCGCCAGAGGGCCCAATTCAAGGCTCAGGGAAGAAATCTATGGAACAACTCAAGGA 1294
QY 1057 GGAGAACTCTCTTTGTTCCCAAGAGGAGGAGTGAAGCTGGAGACCCACATACGTGCCAGC 1116
DB 1295 GGAGAACTCTCTTTGTTCCCAAGAGGAGGAGTGAAGCTGGAGACCCACATACGTGCCAGC 1354
QY 1117 ATCAGCAGCTGCGCGGCTCATTTGGCAAAAGTTGGAAAAACCGTGAACGAGTTTCAGAAATTT 1176
DB 1355 ATCAGCAGCTGCGCGGCTCATTTGGCAAAAGTTGGAAAAACCGTGAACGAGTTTCAGAAATTT 1414
QY 1177 GACGCGCTCAGGTGTAGTACCAAGAGACGACCCCTGATGAGACGACCCAGGTCTAT 1236
DB 1415 GACGCGCTCAGGTGTAGTACCAAGAGACGACCCCTGATGAGACGACCCAGGTCTAT 1474
QY 1237 COTGAAAAATCATCGGACATTTCTATGCCCAGTCAAGTGGCTCAAACGAAAGATCCCGAGCAT 1296
DB 1475 COTGAAAAATCATCGGACATTTCTATGCCCAGTCAAGTGGCTCAAACGAAAGATCCCGAGCAT 1534
QY 1297 COTGCCCCAGGTTAAGCAGACGATCAAGAGGACAGAGTAAACGAGGCCAGGACGAG 1356
DB 1535 COTGCCCCAGGTTAAGCAGACGATCAAGAGGACAGAGTAAACGAGGCCAGGACGAG 1594
QY 1357 GAAGTGACAGCCCTCCCTGTCCTTNGAGTCCAGGACAAACAGCGGCAGAAATCGAGA 1416
DB 1595 GAAGTGACAGCCCTCCCTGTCCTTNGAGTCCAGGACAAACAGCGGCAGAAATCGAGA 1654
QY 1417 GTGTGCTCTCCCGCAGGCTTGAGAAATGAGTGGAAATCCGGGACACNTGGCGCGGCTG 1476
DB 1655 GTGTGCTCTCCCGCAGGCTTGAGAAATGAGTGGAAATCCGGGACACNTGGCGCGGCTG 1714
QY 1477 TAGATCAGTTTGGCCACTTGTATGAGAAAGATGTTCCAGTGAGGAAACCTGATCTNTCA 1536
DB 1715 TAGATCAGTTTGGCCACTTGTATGAGAAAGATGTTCCAGTGAGGAAACCTGATCTNTCA 1774
QY 1537 GCGCCAAAACACCCCAATTTGGCCCAACACTGNTGCCCCCTCGGGGTGTGAGAAATNT 1596
DB 1775 GCGCCAAAACACCCCAATTTGGCCCAACACTGNTGCCCCCTCGGGGTGTGAGAAATNT 1834
QY 1597 AGCGCAAGGCACTTTTAAACGTGGAATGTTTAAAGAGCTCTCCAGGCCCCCAGAGAGG 1656
DB 1835 AGCGCAAGGCACTTTTAAACGTGGAATGTTTAAAGAGCTCTCCAGGCCCCCAGAGAGG 1894
QY 1657 GTGGATCACACTCAGTGGGGAAGAAAAATAAAATTTCTTCCAGGTTTTAAAA 1708
DB 1895 GTGGATCACACTCAGTGGGGAAGAAAAATAAAATTTCTTCCAGGTTTTAAAA 1946

RESULT 3
AAS26148
ID AAS26148 standard; cDNA; 1182 BP.
XX
XX AAS26148;
AC
XX
XX
XX 07-NOV-2001 (first entry)
XX
XX Human cDNA encoding a novel secreted protein, Seq ID 327.
XX
XX Human; immunosuppressive; antiarthritic; ss; antirheumatic; cytostatic;
KW cardant; vasotropic; cerebroprotective; neurotropic; neuroprotective;
KW antibacterial; virucide; fungicide; ophthalmological; vulnerary;
KW secreted protein; rheumatoid arthritis; hyperproliferative disorder;
KW cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
KW cerebral ischaemia; angiogenesis; nervous system disorder;
KW Alzheimer's disease; infection; ocular disorder; corneal infection;
KW wound healing; epithelial cell proliferation; skin ageing; food additive;
KW preservative; antiproliferative.
XX
XX Homo sapiens.
XX
XX WO200155322-A2.
XX
XX 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US0001341.
XX
XX 31-JAN-2000; 2000US-01790658P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214896P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226686P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 13-OCT-2000; 2000US-0239355P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246533P.
PR 08-NOV-2000; 2000US-0246534P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246537P.
PR 08-NOV-2000; 2000US-0246538P.
PR 08-NOV-2000; 2000US-0246539P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.

PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249267P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 08-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251969P.
PR 08-DEC-2000; 2000US-0251989P.
PR 11-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 03-JAN-2001; 2001US-0259678P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
XX WFI; 2001-488783/53.
DR P-PSDB; AAU16161.
XX
PT New nucleic acid molecules encoding 461 human secreted proteins for
PT diagnosing, preventing, treating or ameliorating medical conditions and
PT used as food additives or preservatives.
XX
XX Claim 1; SEQ ID NO 327; 980pp; English.
XX
XX The invention relates to isolated nucleic acid molecules and their
CC encoded secreted proteins. The nucleic acids and proteins are used to
CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
CC in diagnosing a pathological condition or susceptibility to a
CC pathologic condition. Antibodies to the proteins can also be used in
CC alleviating symptoms associated with the disorders and in diagnostic
CC immunoassays e.g. radioimmunoassays or enzyme linked immunosorbant assays
CC (ELISA). Disorders which are diagnosed or treated include autoimmune
CC diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.
CC neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac
CC arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiodysplasia,
CC nervous system disorders e.g. Alzheimer's disease, infections caused by
CC bacteria, viruses and fungi and ocular disorders e.g. corneal infection,
CC and many other disorders listed in the specification. The polypeptides
CC can also be used to aid wound healing and epithelial cell proliferation,
CC to prevent skin aging due to sunburn, to maintain organs before
CC transplantation, for supporting cell culture of primary tissues, to
CC regenerate tissues and in chemotaxis. The polypeptides can also be used
CC as a food additive or preservative to increase or decrease storage
CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,
CC minerals, cofactors and other nutritional components. The present
CC sequence encodes a novel secreted protein of the invention. Note: The
CC sequence data for this patent did not form part of the printed

Query Match 43.4%; Score 741; DB 4; Length 1182;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 991; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 713 TCTCAGCTGATCCCTGGCCCTGAACCTGGCTGCTAGGTCTTTTCCAGCTTCATCCAGC 772
Db 136 TCTCAGCTGATCCCTGGCCCTGAACCTGGCTGCTAGGTCTTTTCCAGCTTCATCCAGC 195
Qy 773 GCATGCCGCGCTCCACAGAGCGTTACTGGGGCTGCTCCCTATAGCTCCTTTATGCGAG 832
Db 196 GCATGCCGCGCTCCACAGAGCGTTACTGGGGCTGCTCCCTATAGCTCCTTTATGCGAG 255
Qy 833 GCTCCCGAGGAGAGGTGTCAGGTGTTATCCCGCCAGCAGTGGCGCCATCATC 892
Db 256 GCTCCCGAGGAGAGGTGTCAGGTGTTATCCCGCCAGCAGTGGCGCCATCATC 315

Qy 893 GGCAAGAGGGGGCAGCAGCATCAAAACAGCTCTCCCGGTTTGGCAGGGCTCCATCAAGATT 952
Db 316 GGCAAGAGGGGGCAGCAGCATCAAAACAGCTCTCCCGGTTTGGCAGGGCTCCATCAAGATT 375
Qy 953 GCACACCCGAAAACACCTGACTCCAAAAGTTCGTATGTTATCATCATCTGACCCGACAG 1012
Db 376 GCACACCCGAAAACACCTGACTCCAAAAGTTCGTATGTTATCATCATCTGACCCGACAG 435
Qy 1013 GCCCAATTTCAAGGCTCAGGGGAAGATCTATGGCAAACTCAAGGAGGAGAACTTCTTGGT 1072
Db 436 GCCCAATTTCAAGGCTCAGGGGAAGATCTATGGCAAACTCAAGGAGGAGAACTTCTTGGT 495
Qy 1073 CCCAAGGAGGAAGTGAAGCTCGAGACCCACATACGTGTGCCAGCATCAGCAGCTGGCCG 1132
Db 496 CCCAAGGAGGAAGTGAAGCTCGAGACCCACATACGTGTGCCAGCATCAGCAGCTGGCCG 555
Qy 1133 GTCATTGCAAGAGTGGAAAAACGGTGAACAGTTCGAGAATTTGACGGCAGCTGAGGTG 1192
Db 556 GTCATTGCAAGAGTGGAAAAACGGTGAACAGTTCGAGAATTTGACGGCAGCTGAGGTG 615
Qy 1193 GTAGTACCAAGAGACAGACCCCTGATGAGAACACACAGGTTCATCGTGAATAATCATCGA 1252
Db 616 GTAGTACCAAGAGACAGACCCCTGATGAGAACACACAGGTTCATCGTGAATAATCATCGA 675
Qy 1253 CATTTCTATGCCAGTCAGATGGCTCAACGGAAGATCCGAGACATCTTGGCCAGGTTAAG 1312
Db 676 CATTTCTATGCCAGTCAGATGGCTCAACGGAAGATCCGAGACATCTTGGCCAGGTTAAG 735
Qy 1313 CAGCAGCATCAGAGGGGACAGAGTAACACAGCCCGAGGACGGAAGTGAACAGCCCT 1372
Db 736 CAGCAGCATCAGAGGGGACAGAGTAACACAGCCCGAGGACGGAAGTGAACAGCCCT 795
Qy 1373 CCTGTCCCTTNGAGTCCAGGACAAACACGGGCGAGAAATCGAGAGTGTGCTCTCCCGGC 1432
Db 796 CCTGTCCCTTNGAGTCCAGGACAAACACGGGCGAGAAATCGAGAGTGTGCTCTCCCGGC 855
Qy 1433 AGGCTGTGAGATGAGTGGGAATCCGGGACACNTGGCGGGCTGTAGATCAGGTTGCC 1492
Db 856 AGGCTGTGAGATGAGTGGGAATCCGGGACACNTGGCGGGCTGTAGATCAGGTTGCC 915
Qy 1493 ACTTGATTGAGAAAGATGTTCCAGTGAAGAAACCTGTATCTTCAGCCCAACACCCACC 1552
Db 916 ACTTGATTGAGAAAGATGTTCCAGTGAAGAAACCTGTATCTTCAGCCCAACACCCACC 975
Qy 1553 CAATTGGCCCAACATGNTGCCCTCGGGTGTGAGAAATNTAGCCAGGACATTTT 1612
Db 976 CAATTGGCCCAACATGNTGCCCTCGGGTGTGAGAAATNTAGCCAGGACATTTT 1035
Qy 1613 AAACGTGATTTGTTTAAAGAGCTCTCCAGGCCCCCAAGAGGTTGATCACACTCAG 1672
Db 1036 AAACGTGATTTGTTTAAAGAGCTCTCCAGGCCCCCAAGAGGTTGATCACACTCAG 1095
Qy 1673 TGGGAAGAAAAATAAAATTTCTTCAGGTTTAAAA 1708
Db 1096 TGGGAAGAAAAATAAAATTTCTTCAGGTTTAAAA 1131

RESULT 4
ABX73489
ID ABX73489 standard; DNA; 1182 BP.
XX
AC ABX73489;
XX
XX 18-MAR-2003 (first entry)
XX
XX Human novel polynucleotide #317.
XX
KW Human; gene; ds; neural disorder; immune system disorder; renal disorder;
KW muscular disorder; respiratory disease; reproductive disorder;
KW gastrointestinal disorder; pulmonary disorder; cardiovascular disorder;
KW hyperproliferative disorder; inflammatory disease; allergic reaction;
KW blood related disorder; cancer; immunosuppressive; antiinflammatory;
KW cardiovascular; nephrotropic; cytostatic; antiallergic; thrombolytic;

KW haemostatic; antiarteriosclerotic.
 XX Homo sapiens.
 OS US2002132753-A1.
 XX 19-SEP-2002.
 XX 17-JAN-2001; 2001US-00764864.
 XX 31-JAN-2000; 2000US-0179065P.
 PR 04-FEB-2000; 2000US-0180628P.
 PR 28-JUN-2000; 2000US-0214886P.
 PR 07-JUL-2000; 2000US-0216647P.
 PR 07-JUL-2000; 2000US-0216880P.
 PR 11-JUL-2000; 2000US-0217487P.
 PR 11-JUL-2000; 2000US-0217496P.
 PR 14-JUL-2000; 2000US-0218293P.
 PR 26-JUL-2000; 2000US-0220963P.
 PR 26-JUL-2000; 2000US-0220964P.
 PR 14-AUG-2000; 2000US-0224518P.
 PR 14-AUG-2000; 2000US-0224519P.
 PR 14-AUG-2000; 2000US-0225267P.
 PR 14-AUG-2000; 2000US-0225268P.
 PR 14-AUG-2000; 2000US-0225270P.
 PR 14-AUG-2000; 2000US-0225447P.
 PR 14-AUG-2000; 2000US-0225757P.
 PR 22-AUG-2000; 2000US-0225758P.
 PR 22-AUG-2000; 2000US-0226868P.
 PR 30-AUG-2000; 2000US-0228924P.
 PR 01-SEP-2000; 2000US-0229287P.
 PR 01-SEP-2000; 2000US-0229289P.
 PR 01-SEP-2000; 2000US-0229343P.
 PR 01-SEP-2000; 2000US-0229344P.
 PR 01-SEP-2000; 2000US-0229345P.
 PR 05-SEP-2000; 2000US-0229509P.
 PR 05-SEP-2000; 2000US-0229513P.
 PR 08-SEP-2000; 2000US-0231413P.
 PR 21-SEP-2000; 2000US-0234223P.
 PR 21-SEP-2000; 2000US-0234274P.
 PR 25-SEP-2000; 2000US-0234997P.
 PR 27-SEP-2000; 2000US-0235834P.
 PR 29-SEP-2000; 2000US-0236327P.
 PR 29-SEP-2000; 2000US-0236367P.
 PR 29-SEP-2000; 2000US-0236369P.
 PR 29-SEP-2000; 2000US-0236370P.
 PR 02-OCT-2000; 2000US-0236802P.
 PR 02-OCT-2000; 2000US-0237037P.
 PR 02-OCT-2000; 2000US-0237038P.
 PR 02-OCT-2000; 2000US-0237039P.
 PR 02-OCT-2000; 2000US-0237040P.
 PR 13-OCT-2000; 2000US-0239935P.
 PR 20-OCT-2000; 2000US-0240960P.
 PR 20-OCT-2000; 2000US-0241785P.
 PR 20-OCT-2000; 2000US-0241809P.
 PR 01-NOV-2000; 2000US-0244617P.
 PR 17-NOV-2000; 2000US-0249299P.
 PR 08-DEC-2000; 2000US-0251856P.
 PR 08-DEC-2000; 2000US-0251868P.
 PR 08-DEC-2000; 2000US-0251869P.
 XX (ROSE/) ROSEN C A.
 PA (RUBE/) RUBEN S M.
 PA (BARA/) BARASH S C.
 XX Rosen CA, Ruben SM, Barash SC;
 PI WPI; 2003-147444/14.
 DR P-PSDB; ABUS5229.
 XX New polypeptides and nucleic acids, useful in gene therapy for treating,
 PT inhibiting or preventing e.g. neural, immune system, muscular,
 PT respiratory, reproductive, gastrointestinal, pulmonary, cardiovascular or

PT renal disorders.
 XX Claim 1; SEQ ID NO 327; 402pp; English.
 XX The invention relates to human novel polypeptides and their associated polynucleotides. The polypeptides and polynucleotides are useful in gene therapy for treating, inhibiting or preventing neural disorders, immune system disorders (e.g. systemic lupus erythematosus, rheumatoid arthritis and multiple sclerosis), muscular disorders, respiratory diseases (e.g. nasal vestibulitis, nasal polyps and sinusitis), reproductive disorders, CC gastrointestinal disorders, pulmonary disorders, cardiovascular disorders, CC (e.g. congenital heart defects, Ebstein's anomaly and hypoplastic left CC heart syndrome), renal disorders (e.g. acute kidney failure and end-stage CC renal disease), hyperproliferative disorders (e.g. Hodgkin's disease and CC leukaemia), inflammatory diseases (e.g. septic shock, Bursitis and CC appendicitis), allergic reactions and conditions (e.g. asthma), blood CC related disorders (e.g. thrombosis, atherosclerosis and myocardial CC infarction) and cancerous diseases. Sequences ABX73173-ABX74167 represent CC human novel polynucleotides of the invention
 XX
 SQ Sequence 1182 BP; 343 A; 310 C; 314 G; 215 T; 0 U; 0 Other;
 Query Match 43.4%; Score 741; DB 7; Length 1182;
 Best Local Similarity 99.5%; Pred. No. 0;
 Matches 991; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 QY 713 TCTCACCTGATCCCTGGCGCTGAACCTGGCTGCTAGTGTCTTTCCAGCTTCATCCAGC 772
 Db 136 TCTCACCTGATCCCTGGCGCTGAACCTGGCTGCTAGTGTCTTTCCAGCTTCATCCAGC 195
 QY 773 GCAGTCCCGCGCTCCCGCAGCGGTTACTGGGGCTGCTCCCTATAGCTCCTTTATGCAG 832
 Db 196 GCAGTCCCGCGCTCCCGCAGCGGTTACTGGGGCTGCTCCCTATAGCTCCTTTATGCAG 255
 QY 833 GCTCCGAGCAGGAGATGGTGCAGGTGTTATCCCGCCAGCAGTGGCGGCATCATC 892
 Db 256 GCTCCGAGCAGGAGATGGTGCAGGTGTTATCCCGCCAGCAGTGGCGGCATCATC 315
 QY 893 GGCAGAGGGGCGCAGCAGCATCAACAGCTCTCCGGTTTCCAGCGCTCCATCAAGATT 952
 Db 316 GGCAGAGGGGCGCAGCAGCATCAACAGCTCTCCGGTTTCCAGCGCTCCATCAAGATT 375
 QY 953 GCACACCCGNAACCTGACTCCAAAGTTCGTATGTTATCATCATCTAGCCGCCAGAG 1012
 Db 376 GCACACCCGNAACCTGACTCCAAAGTTCGTATGTTATCATCATCTAGCCGCCAGAG 435
 QY 1013 GCCCAATTCAAGGCTCAGGGAAGAACTCTATGGCAAACTCAAGGAGGAGAACTTCTTGGT 1072
 Db 436 GCCCAATTCAAGGCTCAGGGAAGAACTCTATGGCAAACTCAAGGAGGAGAACTTCTTGGT 495
 QY 1073 CCCAAGAGGAAGTGAAGCTGGAGACCCATAGTGTGCCAGCATCAGAGCTGGCCGG 1132
 Db 496 CCCAAGAGGAAGTGAAGCTGGAGACCCATAGTGTGCCAGCATCAGAGCTGGCCGG 555
 QY 1133 GTCAATTGGCAAAAGTGAAGAAACGGTGAACGAGTTGAGAAATTTGACGGCAGCTGAGGTG 1192
 Db 556 GTCAATTGGCAAAAGTGAAGAAACGGTGAACGAGTTGAGAAATTTGACGGCAGCTGAGGTG 615
 QY 1193 GTAGTACCAAGAGACCAGACCCCTGTATGAGAACGACCAAGCTCATCGTGAATAATCATCGGA 1252
 Db 616 GTAGTACCAAGAGACCAGACCCCTGTATGAGAACGACCAAGCTCATCGTGAATAATCATCGGA 675
 QY 1253 CATTTCTATGCCAGTCCAGTGGCTCAACGGAAGTCCGAGCATCTGCGCCAGGTTAAG 1312
 Db 676 CATTTCTATGCCAGTCCAGTGGCTCAACGGAAGTCCGAGCATCTGCGCCAGGTTAAG 735
 QY 1313 CAGCAGCATCAGAAGGGACAGAGTAACACAGGCCCCAGCAGCGAGGAAGTGACAGCCCT 1372
 Db 736 CAGCAGCATCAGAAGGGACAGAGTAACACAGGCCCCAGCAGCGAGGAAGTGACAGCCCT 795
 QY 1373 CCCTGTCCCTTNGAGTCCAGGACAAACACGGGGAGAAATCGAGAGTGTCTCTCCCGGC 1432
 Db 796 CCCTGTCCCTTNGAGTCCAGGACAAACACGGGGAGAAATCGAGAGTGTCTCTCCCGGC 855

QY 1433 AGGCTGAGATGAGTGGGATCCGGACACACNTGGCGGGCTGTAGATCAGGTTTGCCC 1492
 Db 856 AGGCTGAGATGAGTGGGATCCGGACACACCTGGCGGGCTGTAGATCAGGTTTGCCC 915
 QY 1493 ACTTGATGAGAAAGATGTTCCAGTGAGGAACCTGTATCTNTCAGCCCCCAACACCCACC 1552
 Db 916 ACTTGATGAGAAAGATGTTCCAGTGAGGAACCTGTATCTNTCAGCCCCCAACACCCACC 975
 QY 1553 CAATTGGCCCAACACTGNTGCCCCCTCGGGGTGTCAGAAATNTTAGGCAAGGCACTTTT 1612
 Db 976 CAATTGGCCCAACACTGTCGCCCTCGGGGTGTCAGAAATNTTAGGCAAGGCACTTTT 1035
 QY 1613 AAACGTGATGTTTAAAGAGCTCTCCAGGCCCCACCAAGAGGGTGATCACACCTCAG 1672
 Db 1036 AAACGTGATGTTTAAAGAGCTCTCCAGGCCCCACCAAGAGGGTGATCACACCTCAG 1095
 QY 1673 TGGGAAGAAAAATAAAATTCCTTCAGGTTTTAAAA 1708
 Db 1096 TGGGAAGAAAAATAAAATTCCTTCAGGTTTTAAAA 1131

RESULT 5

AAK94782

ID AAK94782 standard; cDNA; 2780 BP.

XX AC

XX AAK94782;

XX 06-NOV-2001 (first entry)

XX 06-NOV-2001 (first entry)

XX Human full-length cDNA, SEQ ID NO: 3886.

XX Human; full length cDNA; cDNA synthesis; oligo-capping; ss.

XX Homo sapiens.

XX EPI130094-A2.

XX 05-SEP-2001.

XX 07-JUL-2000; 2000EP-00114089.

XX 08-JUL-1999; 99JP-00194486.

XX 11-JAN-2000; 2000JP-00118774.

XX 02-MAY-2000; 2000JP-00183765.

XX (HELI-) HELIX RES INST.

XX Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;

XX Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;

XX WPI; 2001-524255/58.

XX P-PSDB; AAM93826.

830 Primers useful for synthesizing full length cDNA clones and their use in genetic manipulation.

Claim 8; SEQ ID NO 3886; 1380pp + Sequence Listing; English.

The invention relates to primers for synthesizing full length cDNA clones. 830 cDNA molecules encoding a human protein have been isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA molecules have been determined. Primers for synthesizing the full length cDNA are useful for clarifying the function of the protein encoded by the cDNA. The full length clones were obtained by construction of full length enriched cDNA libraries that were synthesised by the oligo-capping method. The primers enable the production of the full length cDNA easily without any special methods. The present sequence is a full length human cDNA of the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in CD-ROM format directly from EPO

Sequence 2780 BP; 768 A; 681 C; 671 G; 660 T; 0 U; 0 Other;

Query Match 43.4%; Score 741; DB 4; Length 2780;
 Best Local Similarity 99.5%; Pred. No. 0;
 Matches 991; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 713 TCTCACTGATCCCTGGCTGAACCTGGCTGTGTAGTCTTTTCCAGCTTCATCCAGC 772
 Db 421 TCTCACTGATCCCTGGCTGAACCTGGCTGTGTAGTCTTTTCCAGCTTCATCCAGC 480
 QY 773 GCAGTCCCGCCCTCCAGCAGCGTTACTG3GGGTGCTCCCTATAGCTTCCTTTATGCAG 832
 Db 481 GCAGTCCCGCCCTCCAGCAGCGTTACTG3GGGTGCTCCCTATAGCTTCCTTTATGCAG 540
 QY 833 GCTCCGAGCAGAGATGGTGCAGGTGTTATCCCGCCAGGCACTGGGCGCCATCATC 892
 Db 541 GCTCCGAGCAGAGATGGTGCAGGTGTTATCCCGCCAGGCACTGGGCGCCATCATC 600
 QY 893 GGCAAGAGGGGCGAGCAGATCAAAACAGCTCTCCCGGTTTGCAGCGCTCCATCAAGATT 952
 Db 601 GGCAAGAGGGGCGAGCAGATCAAAACAGCTCTCCCGGTTTGCAGCGCTCCATCAAGATT 660
 QY 953 GCACACCGGAAACACCTGACTCCAAAGTTGTTATGTTTATCATCTGGACCGCAGAG 1012
 Db 661 GCACACCGGAAACACCTGACTCCAAAGTTGTTATGTTTATCATCTGGACCGCAGAG 720
 QY 1013 GCCCAATTCAAGGCTCAGGGAAGATCTATGGCAAACTCAAGGAGGAGAACTTCTTTGGT 1072
 Db 721 GCCCAATTCAAGGCTCAGGGAAGATCTATGGCAAACTCAAGGAGGAGAACTTCTTTGGT 780
 QY 1073 CCCAAGGAGAGTGAAGCTGGAGACCAATACGTGTGCCAGCATCAGCAGTGGCCGG 1132
 Db 781 CCCAAGGAGAGTGAAGCTGGAGACCAATACGTGTGCCAGCATCAGCAGTGGCCGG 840
 QY 1133 GTCATTGGCAAGGTGGAAACCGGTGAACGAGTTGCAGAAATTTTACCGCAGCTGAGGTG 1192
 Db 841 GTCATTGGCAAGGTGGAAACCGGTGAACGAGTTGCAGAAATTTTACCGCAGCTGAGGTG 900
 QY 1193 GTAGTACCAAGAGACAGACCCCTGATGAGAACCGACCGAGTTCATCGTGAATAATCATCGGA 1252
 Db 901 GTAGTACCAAGAGACAGACCCCTGATGAGAACCGACCGAGTTCATCGTGAATAATCATCGGA 960
 QY 1253 CATTTCTATGCCAGTCAGATGGCTCAACGGAGATCCGAGACATCCTGGCCAGGTTAAG 1312
 Db 961 CATTTCTATGCCAGTCAGATGGCTCAACGGAGATCCGAGACATCCTGGCCAGGTTAAG 1020
 QY 1313 CAGCAGCATCAGAAGGACAGAGTAAACAGGCCCGCAGGACCGAGGAAGTGACAGCCCT 1372
 Db 1021 CAGCAGCATCAGAAGGACAGAGTAAACAGGCCCGCAGGACCGAGGAAGTGACAGCCCT 1080
 QY 1373 CCTCTCCTTNGAGTCCAGGACAAACACGGGCGAGAAATCGAGAGTGTCTCTCCCGGC 1432
 Db 1081 CCTCTCCTTNGAGTCCAGGACAAACACGGGCGAGAAATCGAGAGTGTCTCTCCCGGC 1140
 QY 1433 AGGCTGAGATGAGTGGGAATCCGGGACACNTGGCGGGCTGTAGATCAGGTTTGCC 1492
 Db 1141 AGGCTGAGATGAGTGGGAATCCGGGACACCTGGCGGGCTGTAGATCAGGTTTGCC 1200
 QY 1493 ACTTGATTGAGAAAGATGTTCCAGTGAGGAACCTGTATCTNTCAGCCCCCAACACCCACC 1552
 Db 1201 ACTTGATTGAGAAAGATGTTCCAGTGAGGAACCTGTATCTCTCAGCCCCCAACACCCACC 1260
 QY 1553 CAATTGGCCCAACACTGNTGCCCTCGGGGTGTCAGAAATNTTAGCCAGGCACTTTT 1612
 Db 1261 CAATTGGCCCAACACTGNTGCCCTCGGGGTGTCAGAAATNTTAGCCAGGCACTTTT 1320
 QY 1613 AAACGTGATGTTTAAAGAGAGCTCTCCAGGCCCCCAACAGAGGGTGATCACACCTCAG 1672
 Db 1321 AAACGTGATGTTTAAAGAGAGCTCTCCAGGCCCCCAACAGAGGGTGATCACACCTCAG 1380
 QY 1673 TGGGAAGAAAAATAAAATTCCTTCAGGTTTTAAAA 1708
 Db 1381 TGGGAAGAAAAATAAAATTCCTTCAGGTTTTAAAA 1416

RESULT 6
AAS26566
ID AAS26566 standard; cDNA; 1129 BP.
XX
AC AAS26566;
XX
DT 07-NOV-2001 (first entry)
XX
DE Human cDNA encoding a novel secreted protein, Seq ID 745.
XX
KW Human; immunosuppressive; antiarthritic; ss; antirheumatic; cytostatic;
KW cardiant; vasotropic; cerebroprotective; nootropic; neuroprotective;
KW antibacterial; virucide; fungicide; ophthalmological; vulnery;
KW secreted protein; rheumatoid arthritis; hyperproliferative disorder;
KW cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
KW cerebral ischaemia; angiogenesis; nervous system disorder;
KW Alzheimer's disease; infection; ocular disorder; corneal infection;
KW wound healing; epithelial cell proliferation; skin ageing; food additive;
KW preservative; antiproliferative.
XX
OS Homo sapiens.
XX
PN WO200155322-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001341.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218230P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 23-AUG-2000; 2000US-0227182P.
PR 30-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234233P.
PR 21-SEP-2000; 2000US-0234234P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 13-OCT-2000; 2000US-0239935P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 01-NOV-2000; 2000US-0241826P.
PR 08-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.

KW cardiovascular; nephrotropic; cytostatic; antiallergic; thrombolytic;
KW haemostatic; antiarteriosclerotic.
XX Homo sapiens.

OS

XX US2002132753-A1.

XX 19-SEP-2002.

XX 17-JAN-2001; 2001US-00764864.

XX 31-JAN-2000; 2000US-0179065P.

XX 04-FEB-2000; 2000US-0180628P.

XX 28-JUN-2000; 2000US-0214886P.

XX 07-JUL-2000; 2000US-0216647P.

XX 07-JUL-2000; 2000US-0216880P.

XX 11-JUL-2000; 2000US-0217487P.

XX 11-JUL-2000; 2000US-0217496P.

XX 14-JUL-2000; 2000US-0218290P.

XX 26-JUL-2000; 2000US-0220963P.

XX 26-JUL-2000; 2000US-0220964P.

XX 14-AUG-2000; 2000US-0224518P.

XX 14-AUG-2000; 2000US-0224519P.

XX 14-AUG-2000; 2000US-0225267P.

XX 14-AUG-2000; 2000US-0225268P.

XX 14-AUG-2000; 2000US-0225270P.

XX 14-AUG-2000; 2000US-0225447P.

XX 14-AUG-2000; 2000US-0225757P.

XX 22-AUG-2000; 2000US-0225758P.

XX 30-AUG-2000; 2000US-0228924P.

XX 01-SEP-2000; 2000US-0229287P.

XX 01-SEP-2000; 2000US-0229343P.

XX 01-SEP-2000; 2000US-0229344P.

XX 01-SEP-2000; 2000US-0229345P.

XX 05-SEP-2000; 2000US-0229509P.

XX 05-SEP-2000; 2000US-0229513P.

XX 08-SEP-2000; 2000US-0231413P.

XX 21-SEP-2000; 2000US-0234223P.

XX 21-SEP-2000; 2000US-0234274P.

XX 25-SEP-2000; 2000US-0234997P.

XX 27-SEP-2000; 2000US-0235834P.

XX 29-SEP-2000; 2000US-0236327P.

XX 29-SEP-2000; 2000US-0236367P.

XX 29-SEP-2000; 2000US-0236368P.

XX 29-SEP-2000; 2000US-0236369P.

XX 29-SEP-2000; 2000US-0236370P.

XX 02-OCT-2000; 2000US-0236802P.

XX 02-OCT-2000; 2000US-0237037P.

XX 02-OCT-2000; 2000US-0237038P.

XX 02-OCT-2000; 2000US-0237039P.

XX 02-OCT-2000; 2000US-0237040P.

XX 13-OCT-2000; 2000US-0239335P.

XX 20-OCT-2000; 2000US-0240960P.

XX 20-OCT-2000; 2000US-0241785P.

XX 01-NOV-2000; 2000US-0241809P.

XX 17-NOV-2000; 2000US-0244617P.

XX 08-DEC-2000; 2000US-0249299P.

XX 08-DEC-2000; 2000US-0251866P.

XX 08-DEC-2000; 2000US-0251868P.

XX 08-DEC-2000; 2000US-0251869P.

XX (ROSE/) ROSEN C A.
PA (RUBE/) RUBEN S M.
PA (BARA/) BARASH S C.

PI Rosen CA, Ruben SM, Barash SC;

XX WPI; 2003-147444/14.

DR P-PSDB; ABUS5647.

XX

PT New polypeptides and nucleic acids, useful in gene therapy for treating,
PT inhibiting or preventing e.g. neural, immune system, muscular.

PT respiratory, reproductive, gastrointestinal, pulmonary, cardiovascular or
PT renal disorders.

XX Claim 1; SEQ ID NO 745; 402pp; English.

XX The invention relates to human novel polypeptides and their associated
CC polynucleotides. The polypeptides and polynucleotides are useful in gene
CC therapy for treating, inhibiting or preventing neural disorders, immune
CC system disorders (e.g. systemic lupus erythematosus, rheumatoid arthritis
CC and multiple sclerosis), muscular disorders, respiratory diseases (e.g.
CC nasal vestibulitis, nasal polyps and sinusitis), reproductive disorders,
CC gastrointestinal disorders, pulmonary disorders, cardiovascular disorders,
CC (e.g. congenital heart defects, Ebstein's anomaly and hypoplastic left
CC heart syndrome), renal disorders (e.g. acute kidney failure and end-stage
CC renal disease), hyperproliferative disorders (e.g. Hodgkin's disease and
CC leukaemia), inflammatory diseases (e.g. septic shock, bursitis and
CC appendicitis), allergic reactions and conditions (e.g. asthma), blood
CC related disorders (e.g. thrombosis, atherosclerosis and myocardial
CC infarction) and cancerous diseases. Sequences ABX73173-ABX74167 represent
CC human novel polynucleotides of the invention

XX Sequence 1129 BP; 318 A; 296 C; 302 G; 210 T; 0 U; 3 Other;

Query Match 34.4%; Score 588; DB 7; Length 1129;

Best Local Similarity 99.2%; Pred. No. 6.5e-285;

Matches 988; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 713 TCTCAGCTGATCCCTGGCTGCAACCTGGCTGCTGCTAGGTCTTTTCCAGCTTCATCCAGC 772

DB 103 TCTCAGCTGATCCCTGGCTGCAACCTGGCTGCTGCTAGGTCTTTTCCAGCTTCATCCAGC 162

QY 773 GCAGTCCGCGCTCCAGCAGCGTTACTGGGGCTGCTCCCTATAGCTCCTTTATGCGAG 832

DB 163 GCAGTCCGCGCTCCAGCAGCGTTACTGGGGCTGCTCCCTATAGCTCCTTTATGCGAG 222

QY 833 GCTCCGAGCAGGAGATGGTGCAGTGTTCATCCCGCCAGCAGTGGCGGCATCATC 892

DB 223 GCTCCGAGCAGGAGATGGTGCAGTGTTCATCCCGCCAGCAGTGGCGGCATCATC 282

QY 893 GGCAGAGAGGCGCAGCAGCATCAACAGCTCTCCGGTTTCCAGCGCTCCATCAAGATT 952

DB 283 GGCAGAGAGGCGCAGCAGCATCAACAGCTCTCCGGTTTCCAGCGCTCCATCAAGATT 342

QY 953 GCACACCCGAAACCTGACTCCAAAGTTGATGTTGATGTTATCATCTGACCGCCAGAG 1012

DB 343 GCACACCCGAAACCTGACTCCAAAGTTGATGTTGATGTTATCATCTGACCGCCAGAG 402

QY 1013 GCCCAATTCAAGGCTCAGGGAAGAATCTATGGCAAACTCAAGGAGGAGAACTCTTTTGGT 1072

DB 403 GCCCAATTCAAGGCTCAGGGAAGAATCTATGGCAAACTCAAGGAGGAGAACTCTTTTGGT 462

QY 1073 CCCAAGAGGAGAGTGAAGCTGGAGACCCATAGCTGTGCGCAGCATCAGAGTGGCCGG 1132

DB 463 CCCAAGAGGAGAGTGAAGCTGGAGACCCATAGCTGTGCGCAGCATCAGAGTGGCCGG 522

QY 1133 GTCAATTGGCAAAAGTGGAAAAACGGTGAACGAGTTGAGAAATTTGACGGCAGCTGAGGTG 1192

DB 523 GTCAATTGGCAAAAGTGGAAAAACGGTGAACGAGTTGAGAAATTTGACGGCAGCTGAGGTG 582

QY 1193 GTAGTACCAAGAGACCCAGACCCCTGTATGAGAACGACAGTCTATCGTGAATAATCATCGGA 1252

DB 583 GTAGTACCAAGAGACCCAGACCCCTGTATGAGAACGACAGTCTATCGTGAATAATCATCGGA 642

QY 1253 CATTTCTATGCCAGTCAAGTGGCTCAACGGAAGATCCGAGACATCTTGGCCAGGTTAAG 1312

DB 643 CATTTCTATGCCAGTCAAGTGGCTCAACGGAAGATCCGAGACATCTTGGCCAGGTTAAG 702

QY 1313 CAGCAGCATCAGAGGAGACAGAGTAACAGGCCAGCGCCAGCGAGGAGTGAACAGCCCT 1372

DB 703 CAGCAGCATCAGAGGAGACAGAGTAACAGGCCAGCGCCAGCGAGGAGTGAACAGCCCT 762

QY 1373 CCCTGTCCCTTNGAGTCCAGGACCAACAGGGCAGAAATCGAGAGTGTCTCTCCCGGC 1432

Db 763 CCTGTCCCTTCAGTCCAGGCAACAAACGGGAGAAATCGAGAGTGTGCTCTCTCCCGGC 822
Qy 1433 AGGCTGAGAAATGAGTGGGAATCCGGGACACNCTGGGCGGCTGTGATGACAGTTTGCCC 1492
Db 823 AGGCTGAGAAATGAGTGGGAATCCGGGACACNCTGGGCGGCTGTGATGACAGTTTGCCC 882
Qy 1493 ACTTGATTGAGAAAGATGTTCCAGTGAAGAACCTGATCTNTCAGCCCCCAACACCCACC 1552
Db 883 ACTTGATTGAGAAAGATGTTCCAGTGAAGAACCTGATCTCTCAGCCCCCAACACCCACC 942
Qy 1553 CAATTGGCCCAACACTGTNTGCCCTCGGGGTCTCAGAAATNTAGCGAAGCACTTTT 1612
Db 943 CAATTGGCCCAACACTGTCTGCCCTCGGGGTCTCAGAAATNTAGCGAAGCACTTTT 1002
Qy 1613 AAACGTGGATTGTTTAAAGAGCTCTCCAGGCCCCACCAAGAGGGTGGATCACACCTCAG 1672
Db 1003 AAACGTGGATTGTTTAAAGAGCTCTCCAGGCCCCACCAAGAGGGTGGATCACACCTCAG 1062
Qy 1673 TGGGAAGAAAAATAAAATTTCCCTTCAGGTTTTTAAA 1708
Db 1063 TGGGAAGAAAAATAAAATTTCCCTTCAGGTTTTTAAA 1098

RESULT 8

AAK91969
ID AAK91969 standard; cDNA; 833 BP.
XX
AC AAK91969;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human cDNA 5'-end sequence, SEQ ID NO: 429.
XX
KW Human; full length cDNA; cDNA synthesis; oligo-capping; ss.
XX
OS Homo sapiens.
XX
FN EP130094-A2.
XX
PD 05-SEP-2001.
XX
PF 07-JUL-2000; 2000EP-00114089.
XX
PR 08-JUL-1999; 99JP-00194486.
PR 11-JAN-2000; 2000JP-00118774.
PR 02-MAY-2000; 2000JP-00183765.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX

WPI; 2001-524255/58.
XX
XX 830 Primers useful for synthesizing full length cDNA clones and their use
in genetic manipulation.
XX
PS Claim 2; SEQ ID NO 429; 1380pp + Sequence Listing; English.

XX The invention relates to primers for synthesizing full length cDNA
clones. 830 cDNA molecules encoding a human protein have been isolated
and nucleotide sequences of 5'- and 3'-ends of the cDNA molecules have
been determined. Primers for synthesizing the full length cDNA are useful
for clarifying the function of the protein encoded by the cDNA. The full
length clones were obtained by construction of full length enriched cDNA
libraries that were synthesised by the oligo-capping method. The primers
enable the production of the full length cDNA easily without any special
methods. The present sequence is the nucleotide sequence of the 5'-end of
a cDNA provided in the invention. Note: The sequence data for this patent
did not form part of the printed specification, but was obtained in CD-
ROM format directly from EPO
XX
XX Sequence 833 BP; 223 A; 225 C; 215 G; 166 T; 0 U; 4 Other;

Query Match 24-2k; Score 414; DB 4; Length 833;
Best Local Similarity 100.0%; Pred. No. 2.1e-197; Indels 0; Gaps 0;
Matches 414; Conservative 0; Mismatches 0;
Qy 299 ACAAAACAGACCCAGTCCAGATAGAGTGCATAGGAAGAGAAACGCGAGGTGCAGCTGAA 358
Db 1 ACAAAACAGACCCAGTCCAGATAGAGTGCATAGGAAGAGAAACGCGAGGTGCAGCTGAA 60
Qy 359 AAGCCATCAGTGTGCTACTCCACCCCTGAGGGCTGCTCCTCGCTGTGTAAGATGATCTTG 418
Db 61 AAGCCATCAGTGTGCTACTCCACCCCTGAGGGCTGCTCCTCGCTGTGTAAGATGATCTTG 120
Qy 419 GAGATTATGCATAAAGAGGCTAAGACACCAAAACGGCTGACGAGGTTCCCTCGAAGATC 478
Db 121 GAGATTATGCATAAAGAGGCTAAGACACCAAAACGGCTGACGAGGTTCCCTCGAAGATC 180
Qy 479 GTGGCCCATATAACTTTGTAGGGGTCTCTATTGGCAAGAGGACGGAACCTGAGGAAG 538
Db 181 GTGGCCCATATAACTTTGTAGGGGTCTCTATTGGCAAGAGGACGGAACCTGAGGAAG 240
Qy 539 GTAGAGCAAGATACCGAGACAAAATACCATCTCCTCGTTGCAAGACCTTACCTTTAC 598
Db 241 GTAGAGCAAGATACCGAGACAAAATACCATCTCCTCGTTGCAAGACCTTACCTTTAC 300
Qy 599 AACCTTGAGAGACCATCACTGTGAAGGGGCCCATCGAGAAATTTGTCAGGGCCGAGCAG 658
Db 301 AACCTTGAGAGACCATCACTGTGAAGGGGCCCATCGAGAAATTTGTCAGGGCCGAGCAG 360
Qy 659 GAAATATGAAGAAAGTTTCGGAGGCCCTATGAGATGATGTGGTGCATGAGC 712
Db 361 GAAATATGAAGAAAGTTTCGGAGGCCCTATGAGATGATGTGGTGCATGAGC 414

RESULT 9

AAK93655
ID AAK93655 standard; cDNA; 833 BP.
XX
AC AAK93655;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human cDNA clone representative sequence, SEQ ID NO: 2115.
XX
KW Human; full length cDNA; cDNA synthesis; oligo-capping; ss.
XX
OS Homo sapiens.
XX
FN EP130094-A2.
XX
PD 05-SEP-2001.
XX
PF 07-JUL-2000; 2000EP-00114089.
XX
PR 08-JUL-1999; 99JP-00194486.
PR 11-JAN-2000; 2000JP-00118774.
PR 02-MAY-2000; 2000JP-00183765.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX

WPI; 2001-524255/58.
XX
XX 830 Primers useful for synthesizing full length cDNA clones and their use
in genetic manipulation.

PS Example 11; SEQ ID NO 2115; 1380pp + Sequence Listing; English.

XX The invention relates to primers for synthesising full length cDNA
clones. 830 cDNA molecules encoding a human protein have been isolated
and nucleotide sequences of 5'- and 3'-ends of the cDNA molecules have

DR WPI; 2002-257383/30.

XX New oligonucleotide libraries comprising oligonucleotides which

PT selectively hybridize to mRNAs transcribed from a transcription unit of a

PT genome, useful for detecting tissue-, pathology-, and developmental-

PT specific genes.

XX

PS Example 1; SEQ ID NO 31873; 47pp; English.

XX

CC The present invention describes oligonucleotide libraries for detecting

CC messenger RNAs that populate a (sub-)transcriptome, where the (sub-

CC)transcriptome comprises messenger RNAs transcribed from multiple

CC transcription units that populate a genome. The library comprises several

CC oligonucleotides, each capable of hybridising selectively to a set of

CC messenger RNAs transcribed from a given transcription unit of the genome,

CC which encodes one or more messenger RNA splice variants. The

CC oligonucleotide libraries are useful for detecting mRNAs from a

CC biological sample, in expression profiling studies, in qualitatively or

CC quantitatively characterising the corresponding transcriptome, and in

CC detecting RNA transcripts and splice variants of human or animal

CC transcripts. The libraries may also be used as specialised mini

CC libraries to detect transcripts of a sub-transcriptome under a particular

CC biological or pathological state, and so allowing the detection of tissue

CC - and pathology-specific genes such as those genes only expressed in

CC specific tissue under a specific pathological condition; to detect

CC variants of a transcriptome of a patient suffering from a particular

CC developmental specific genes; and to detect RNA transcripts and splice

CC variants of a transcriptome of a patient suffering from a particular

CC disorder. ABN27253 to ABN59589 represent oligonucleotide sequences from

CC rats, humans and mice, which are used in the exemplification of the

CC present invention. N.B. The sequence data for this patent did not form

CC part of the printed specification, but was obtained in electronic format

CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX

SQ Sequence 60 BP; 15 A; 12 C; 18 G; 15 T; 0 U; 0 Other;

Query Match 3.5%; Score 60; DB 6; Length 60;

Best Local Similarity 100.0%; Pred. No. 2.3e-19;

Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1470 CGGCTGTAGATCAGGTTGCCCACTTGATTGAGAAAGATGTTCCAGTGAGGAACCTGA 1529

DB 1 CGGCTGTAGATCAGGTTGCCCACTTGATTGAGAAAGATGTTCCAGTGAGGAACCTGA 60

RESULT 12

ABN58997

ID ABN58997 standard; DNA; 60 BP.

XX

AC ABN58997;

XX

XX 15-JUL-2002 (first entry)

XX

DE Human spliced transcript detection oligonucleotide SEQ ID NO:31745.

XX

XX Human; mouse; rat; splice transcript; detection; RNA transcript;

KW splice variant; transcriptome; oligonucleotide library; ss.

KW

OS Homo sapiens.

XX

XX WO200210449-A2.

PN

XX

PD 07-FEB-2002.

XX

PF 20-JUL-2001; 2001WO-IB001903.

XX

XX 28-JUL-2000; 2000US-0221607P.

PR

PR 02-MAY-2001; 2001US-0287724P.

XX

XX (COMP-) COMPUGEN INC.

PA

XX Shoshan A, Wasserman A, Mintz E, Mintz L, Faigler S;

PI

XX WPI; 2002-257383/30.

DR

XX New oligonucleotide libraries comprising oligonucleotides which

PT selectively hybridize to mRNAs transcribed from a transcription unit of a

PT genome, useful for detecting tissue-, pathology-, and developmental-

PT specific genes.

XX

PS Example 1; SEQ ID NO 31745; 47pp; English.

XX

CC The present invention describes oligonucleotide libraries for detecting

CC messenger RNAs that populate a (sub-)transcriptome, where the (sub-

CC)transcriptome comprises messenger RNAs transcribed from multiple

CC transcription units that populate a genome. The library comprises several

CC oligonucleotides, each capable of hybridising selectively to a set of

CC messenger RNAs transcribed from a given transcription unit of the genome,

CC which encodes one or more messenger RNA splice variants. The

CC oligonucleotide libraries are useful for detecting mRNAs from a

CC biological sample, in expression profiling studies, in qualitatively or

CC quantitatively characterising the corresponding transcriptome, and in

CC detecting RNA transcripts and splice variants of human or animal

CC transcripts. The libraries may also be used as specialised mini

CC libraries to detect transcripts of a sub-transcriptome under a particular

CC biological or pathological state, and so allowing the detection of tissue

CC - and pathology-specific genes such as those genes only expressed in

CC specific tissue under a specific pathological condition; to detect

CC variants of a transcriptome of a patient suffering from a particular

CC developmental specific genes; and to detect RNA transcripts and splice

CC variants of a transcriptome of a patient suffering from a particular

CC disorder. ABN27253 to ABN59589 represent oligonucleotide sequences from

CC rats, humans and mice, which are used in the exemplification of the

CC present invention. N.B. The sequence data for this patent did not form

CC part of the printed specification, but was obtained in electronic format

CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX

SQ Sequence 60 BP; 15 A; 12 C; 18 G; 15 T; 0 U; 0 Other;

Query Match 3.5%; Score 60; DB 6; Length 60;

Best Local Similarity 100.0%; Pred. No. 2.3e-19;

Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1470 CGGCTGTAGATCAGGTTGCCCACTTGATTGAGAAAGATGTTCCAGTGAGGAACCTGA 1529

DB 1 CGGCTGTAGATCAGGTTGCCCACTTGATTGAGAAAGATGTTCCAGTGAGGAACCTGA 60

RESULT 13

ABZ02948

ID ABZ02948 standard; DNA; 50 BP.

XX

AC ABZ02948;

XX

XX 09-JAN-2003 (first entry)

XX

DE Human leukocyte gene expression profiling probe SEQ ID NO 2939.

XX

XX T7; leukocyte; gene expression profiling; allograft rejection;

KW atherosclerosis; congestive heart failure; systemic lupus erythematosus;

KW rheumatoid arthritis; osteoarthritis; cytomegalovirus; infection; probe;

KW ss.

XX

XX Homo sapiens.

OS

XX

XX WO200257414-A2.

PN

XX

PD 25-JUL-2002.

XX

XX 22-OCT-2001; 2001WO-US047856.

PF

XX 20-OCT-2000; 2000US-0241994P.

PR

PR 08-JUN-2001; 2001US-0296764P.

XX

XX (BIOC-) BIOCARDIA INC.

PA

XX Wohlgenuth J, Fry K, Matcuk G, Altman P, Prentice J, Phillips J;

PI Ly N, Woodward R, Quettermus T, Johnson F;

PI

XX WPI; 2002-636525/68.
 XX
 XX
 PT New system for leukocyte expression profiling, diagnosing a disease, or
 PT monitoring (the rate of) progression of a disease, e.g. atherosclerosis
 PT or congestive heart failure, comprises diagnostic oligonucleotides.
 XX
 PS Claim 1; Page 421; Opp; English.
 XX
 XX The invention relates to a system for detecting gene expression, which
 CC comprises one or two isolated DNA molecules that detect expression of a
 CC gene, where the gene corresponds to any of 8143 oligonucleotides
 CC (ABZ0010-ABZ08152) each having 50 base pairs (bp). The system is useful
 CC for leukocyte expression profiling. It is particularly useful for
 CC diagnosing a disease, monitoring (rate of) progression of a disease,
 CC predicting therapeutic outcome, determining prognosis for a patient,
 CC predicting disease complications in an individual or monitoring response
 CC to treatment in an individual. The diseases include cardiac allograft
 CC rejection, kidney allograft rejection, liver allograft rejection,
 CC atherosclerosis, congestive heart failure, systemic lupus erythematosus,
 CC rheumatoid arthritis, osteoarthritis or cytomegalovirus infection
 XX
 SQ Sequence 50 BP; 18 A; 8 C; 13 G; 11 T; 0 U; 0 Other;
 Query Match 2.9%; Score 50; DB 6; Length 50;
 Best Local Similarity 100.0%; Pred. No. 2.5e-14;
 Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1652 AGAGGGTGGATCACCTCAGTGGAGAGAAATAAATTTCTTCAGGT 1701
 Db 1 AGAGGGTGGATCACCTCAGTGGAGAGAAATAAATTTCTTCAGGT 50
 RESULT 14
 AAZ10617
 ID AAZ10617 standard; cDNA; 2224 BP.
 AC AAZ10617;
 XX
 XX 17-NOV-1999 (first entry)
 DE cDNA encoding a murine c-myc coding region determinant binding protein.
 XX
 XX c-myc coding region determinant binding protein; CRD-BP; tumor; c-myc;
 KW endonucleolytic attack; half-life; breast cancer; colon cancer;
 KW pancreatic cancer; ss.
 XX
 OS Mus musculus.
 XX
 FH Key Location/Qualifiers
 CD 131..1864
 FT /*tag= a
 FT
 XX
 PN WO9946594-A2.
 XX
 PD 16-SEP-1999.
 XX
 XX 05-MAR-1999; 99WO-US0004897.
 XX
 XX 09-MAR-1998; 98US-0077372P.
 XX
 PA (WISC) WISCONSIN ALUMNI RES FOUND.
 XX
 PI Ross J;
 XX
 XX WPI; 1999-551506/46.
 DR P-PSDB; AAY30649.
 DR
 DR Diagnosing presence or absence of a tumor in a human by examining c-myc
 PT coding region determinant-binding protein.
 PT
 PS Example; Fig 1A-D; 79pp; English.
 PS
 XX

CC The present sequence encodes a murine c-myc coding region determinant
 CC binding protein (CRD-BP). The presence or absence of a tumor can be
 CC determined by determining the levels of CRD-BP present in the suspect
 CC tissue, where the CRD-BP shields c-myc RNA from endonucleolytic attack
 CC and so prolongs its half-life. The methods are used for diagnosing
 CC presence or absence of a tumor in a human, especially breast, colon and
 CC pancreatic cancer. They are also used to inhibit cancer cell growth
 XX
 SQ Sequence 2224 BP; 580 A; 609 C; 618 G; 417 T; 0 U; 0 Other;
 Query Match 1.8%; Score 31; DB 2; Length 2224;
 Best Local Similarity 100.0%; Pred. No. 9.5e-05;
 Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 66 TGGAGAACCATGCCCTGAAGGTCCTCTCTACAT 96
 Db 561 TGGAGAACCATGCCCTGAAGGTCCTCTCTACAT 591
 RESULT 15
 AAZ10625/c
 ID AAZ10625 standard; DNA; 24 BP.
 AC AAZ10625;
 XX
 XX 17-NOV-1999 (first entry)
 DE PCR primer used to amplify murine CRD-BP cDNA.
 XX
 XX c-myc coding region determinant binding protein; CRD-BP; tumor; c-myc;
 KW endonucleolytic attack; half-life; breast cancer; colon cancer;
 KW pancreatic cancer; PCR primer; ss.
 XX
 OS Synthetic.
 OS Mus musculus.
 XX
 PN WO9946594-A2.
 XX
 PD 16-SEP-1999.
 XX
 XX 05-MAR-1999; 99WO-US0004897.
 XX
 XX 09-MAR-1998; 98US-0077372P.
 XX
 PA (WISC) WISCONSIN ALUMNI RES FOUND.
 XX
 PI Ross J;
 XX
 XX WPI; 1999-551506/46.
 DR
 DR Diagnosing presence or absence of a tumor in a human by examining c-myc
 PT coding region determinant-binding protein.
 PT
 PS Example; Page 27; 79pp; English.
 XX
 XX The present PCR primer was used to amplify cDNA encoding murine c-myc
 CC coding region determinant binding protein (CRD-BP). The presence or
 CC absence of a tumor can be determined by determining the levels of CRD-BP
 CC present in the suspect tissue, where the CRD-BP shields c-myc RNA from
 CC endonucleolytic attack and so prolongs its half-life. The methods are
 CC used for diagnosing presence or absence of a tumor in a human, especially
 CC breast, colon and pancreatic cancer. They are also used to inhibit cancer
 CC cell growth
 XX
 SQ Sequence 24 BP; 3 A; 8 C; 5 G; 8 T; 0 U; 0 Other;
 Query Match 1.4%; Score 24; DB 2; Length 24;
 Best Local Similarity 100.0%; Pred. No. 0.29;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 508 CATTGGCAGGAGGACGGAACCT 531
 Db 24 CATTGGCAGGAGGACGGAACCT 1

Search completed: July 13, 2004, 23:52:48
Job time : 704 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 13, 2004, 23:20:54 ; Search time 4590 seconds

(without alignments)
11112.114 Million cell updates/sec

Title: US-09-270-437D-5

Perfect score: 1708

Sequence: 1 agggagctgcgcacccgc.....atttccttcagggttttaaaa 1708

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 27513289 seqs, 14931090276 residues

Word size : 0

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

BST.*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hcc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hcc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	ID	Description
1	676	39.6	1042 12	BM928196
2	657	38.5	1085 13	BU190334
3	606	35.5	955 13	BQ651665
4	584	34.2	950 13	BQ648383

545	31.9	983	13	BO643920	BO643920	AGENCOURT
526	30.8	831	12	BG115319	602316274	
515	30.2	919	13	EX327672	EX327672	EX327672
499	29.2	896	10	BF984962	602308484	
474	27.8	1042	12	BM561057	AGENCOURT	
426	24.9	1061	13	BQ647561	AGENCOURT	
419	24.5	712	13	BQ225582	AGENCOURT	
416	24.5	1061	13	BQ225582	AGENCOURT	
412	24.4	953	13	BO647360	AGENCOURT	
382	22.4	884	13	BU598549	AGENCOURT	
351	17.3	761	9	AW003366	wq65h07.x	
291	17.0	568	14	CB146278	K-EST0201	
290	17.0	545	13	EX089485	EX089485	
288	16.9	1173	12	BM467135	AGENCOURT	
272	15.9	982	13	BO652180	AGENCOURT	
268	15.7	978	12	BG115593	602317291	
246	14.4	764	12	BM006191	603614013	
232	13.6	929	13	BO652586	AGENCOURT	
231	13.5	486	9	AL704124	DKFp6868	
231	13.5	891	13	EX372273	EX372273	
219	12.8	484	9	AI337147	AI337147	qx83c08.x
213	12.5	550	9	AA978341	AA978341	Qx40a04.s
213	12.5	896	13	BO647425	AGENCOURT	
194	11.4	896	13	BO647425	AGENCOURT	
172	10.1	517	10	BE550506	7a27f01.x	
167	9.8	1201	13	EX401619	EX401619	EX401619
158	9.3	656	10	AW303505	xv19h09.x	
151	8.8	478	9	AA196977	AA196977	zq60h02.s
138	8.1	257	10	BE466868	hzz28a05.x	
135	7.9	249	9	AA196774	AA196774	zq60f02.f
131	7.7	1201	13	EX333010	EX333010	
121	7.1	393	28	AQ096481	HS_3029.B	
121	7.1	628	29	AG064887	Pan_tro91	
113	6.6	265	9	AI962864	wq42f07.x	
113	6.6	317	10	AW206272	UI-H-B11-	
113	6.6	461	9	AA196759	AA196759	zq60f02.s
110	6.4	258	9	AI989990	ws28f09.x	
109	6.4	257	9	AI335528	qt24b06.x	
109	6.4	257	9	AI571822	wb41h04.x	
109	6.4	257	10	BE466712	hzz4e03.x	
109	6.4	258	10	AW196265	xm31e06.x	
109	6.4	258	10	BE550313	7a21h08.x	
109	6.4	369	10	BF115470	7m89a03.x	

ALIGNMENTS

RESULT 1
BM928196
LOCUS
DEFINITION
AGENCOURT_6715416 NIH_MGC_100 Homo sapiens CDNA clone IMAGE:5797961
5', mRNA sequence.
ACCESSION
BM928196
VERSION
BM928196.1
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 1042)
NIH-MGC http://mgc.mci.nih.gov/.
AUTHORS
National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE
Unpublished (1999)
JOURNAL
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: CGAP (Stanford)
CDNA Library Preparation: Rubin Laboratory
DNA Sequencing Arrayed by: The I.M.A.G.E. Consortium (ILNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ILNL at:
http://image.llnl.gov
Place: LLCM2025 row: c column: 18
High quality sequence stop: 662.

FEATURES	Location/Qualifiers	source
LOCUS	1..1042	
DEFINITION	/organism="Homo sapiens"	
	/mol_type="mRNA"	
	/db_xref="taxon:9606"	
ACCESSION	/clone="IMAGE:5797961"	
VERSION	/tissue_type="hepatocellular carcinoma, cell line"	
KEYWORDS	/lab_host="DH10B (phage-resistant)"	
SOURCE	/clone_lib="NIH_MGC_100"	
ORGANISM	/note="Organ: liver; Vector: pOTB7; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."	
ORIGIN		
	Query Match 39.6%; Score 676; DB 12; Length 1042;	
	Best Local Similarity 100.0%; Pred. No. 0;	
	Matches 676; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	37 AGCCATCATGAAGCTGAATGGCCACCATGTGGAGAACCATGCCCTGGAAGGTCTCCTACAT	96
DB	12 AGCCATCATGAAGCTGAATGGCCACCATGTGGAGAACCATGCCCTGGAAGGTCTCCTACAT	71
QY	97 CCCGATGACGATAGACACAGGACCTGAGAACTGGGCGCGAGGGGCTTTGGCTCTCG	156
DB	72 CCCGATGACGATAGACACAGGACCTGAGAACTGGGCGCGAGGGGCTTTGGCTCTCG	131
QY	157 GGGTCAGCCCGCCAGGGCTCACTGTGCGAGCGGGGGCCCCAGCCAGCAGCAGCAGT	216
DB	132 GGGTCAGCCCGCCAGGGCTCACTGTGCGAGCGGGGGCCCCAGCCAGCAGCAGCAGT	191
QY	217 GGACATCCCCCTTCGGCTCCTGGTGCACCACCATGTGGGTGCCATTATTTGCCAAGGA	276
DB	192 GGACATCCCCCTTCGGCTCCTGGTGCACCACCATGTGGGTGCCATTATTTGCCAAGGA	251
QY	277 GGGGGCCACATCCGACATCATCAAAACAGACCCAGTCCAGATAGACGTGCATAGAA	336
DB	252 GGGGGCCACATCCGACATCATCAAAACAGACCCAGTCCAGATAGACGTGCATAGAA	311
QY	337 GGAGAACGCGAGGTGCAGCTGAAAAAGCCATCAGTGTGCATCCACCCCTGAGGGCTGCTC	396
DB	312 GGAGAACGCGAGGTGCAGCTGAAAAAGCCATCAGTGTGCATCCACCCCTGAGGGCTGCTC	371
QY	397 CTCGCTTTAGATGATCTTGGAGATTATGCTAAGAGGCTAAGACACCAAAACGGC	456
DB	372 CTCGCTTTAGATGATCTTGGAGATTATGCTAAGAGGCTAAGACACCAAAACGGC	431
QY	457 TGACAGAGTTCCCTTGAGATCCTGGCCCATAACTTTGTAGGGGTCTCATTTGCCAA	516
DB	432 TGACAGAGTTCCCTTGAGATCCTGGCCCATAACTTTGTAGGGGTCTCATTTGCCAA	491
QY	517 GGAAGACGGAACCTGAAGAGGTAGACAAATACCGAGACAAATAACCAATCTCCTC	576
DB	492 GGAAGACGGAACCTGAAGAGGTAGACAAATACCGAGACAAATAACCAATCTCCTC	551
QY	577 GTTGCAAGACCTTACCCTTTTCAACCCCTGAGAGGACCATCACTGTGAAGGGGGCCATCGA	636
DB	552 GTTGCAAGACCTTACCCTTTTCAACCCCTGAGAGGACCATCACTGTGAAGGGGGCCATCGA	611
QY	637 GAAATGTTTCAGGGCGGACGAGAAATAATGAAGAAAGTTCCGGAGGCGCTATGAGAAATGA	696
DB	612 GAAATGTTTCAGGGCGGACGAGAAATAATGAAGAAAGTTCCGGAGGCGCTATGAGAAATGA	671
QY	697 TGTGGCTGCCATGAGC	712
DB	672 TGTGGCTGCCATGAGC	687

Db 332 GAAAAAGCCATCAGTGTGCACTCCACCCCTGAGGGCTGCTCCTCGGCTTGAAGATGATC 391
Qy 416 TTGAGAGATTATGCTAAAGAGGCTAAGGACACCAAAACGGCTGACGAGGTTCCCTCGAAG 475
Db 392 TTGAGAGATTATGCTAAAGAGGCTAAGGACACCAAAACGGCTGACGAGGTTCCCTCGAAG 451
Qy 476 ATCTGCGCCCATATAACTTTGTAGGGCGTCTCATTTGGCAAGGAAGACCGAAACCTTGAAG 535
Db 452 ATCTGCGCCCATATAACTTTGTAGGGCGTCTCATTTGGCAAGGAAGACCGAAACCTTGAAG 511
Qy 536 AAGTGAGCAAGATACCGAGACAAAATACCAATCTCTCGTTGCAAGACCTTACCTTT 595
Db 512 AAGTGAGCAAGATACCGAGACAAAATACCAATCTCTCGTTGCAAGACCTTACCTTT 571
Qy 596 TACAACCTGAGAGGACCATCAGTGTGAAGGGGSCCATCGAGAAATTTTCAGGGCGGAG 655
Db 572 TACAACCTGAGAGGACCATCAGTGTGAAGGGGSCCATCGAGAAATTTTCAGGGCGGAG 631
Qy 656 CAGGAATAATGAAGAAAGTTTCGGAGGCTTATGAGAAATGATGTGGCTGCCATGAGC 712
Db 632 CAGGAATAATGAAGAAAGTTTCGGAGGCTTATGAGAAATGATGTGGCTGCCATGAGC 688

RESULT 3
B0651665 955 bp mRNA linear EST 15-JUL-2002
LOCUS AGENCOURT 8297761 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6269842
DEFINITION 5', mRNA sequence.
ACCESSION B0651665
VERSION B0651665.1 GI:21775837
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 955)
AUTHORS NIH-MGC http://mgs.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: CGAP (Stanford)
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LNCM2444 row: a column: 11
High quality sequence stop: 588.
Location/Qualifiers
1..955
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6269842"
/tissue_type="hepatocellular carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_100"
/note="Organ: liver; Vector: pOTB7; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGGAG(G). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-cDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH_MGC
Library."

FEATURES
source
1..955
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6269842"
/tissue_type="hepatocellular carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_100"
/note="Organ: liver; Vector: pOTB7; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGGAG(G). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-cDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH_MGC
Library."

Query Match 35.5%; Score 606; DB 13; Length 955;
Best Local Similarity 100.0%; Pred. No. 2.7e-301;
Matches 606; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

ORIGIN

Query Match 35.5%; Score 606; DB 13; Length 955;
Best Local Similarity 100.0%; Pred. No. 2.7e-301;
Matches 606; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 37 AGCCATCATGAGCTGAATGGCCACACAGTTGGAGAACCATGCCCTGAAGTCTCTACAT 96
Db 12 AGCCATCATGAGCTGAATGGCCACACAGTTGGAGAACCATGCCCTGAAGTCTCTACAT 71
Qy 97 CCCCGATGAGCAGATAGCACAGGACCTGAGAAATGGGCGCCGAGGGGCTTTGGTCTCTCG 156
Db 72 CCCCGATGAGCAGATAGCACAGGACCTGAGAAATGGGCGCCGAGGGGCTTTGGTCTCTCG 131
Qy 157 GGGTCAGCCCCCGCAGGGCTCACCTGTGGCAGCGGGGGCCCCAGCAAGCAGCAGCAAGT 216
Db 132 GGGTCAGCCCCCGCAGGGCTCACCTGTGGCAGCGGGGGCCCCAGCAAGCAGCAGCAAGT 191
Qy 217 GGACATATCCCTTCCGCTCCTGGTCCCAACCCAGTATGTGGGTGCCATTTATTGGCAAGA 276
Db 192 GGACATATCCCTTCCGCTCCTGGTCCCAACCCAGTATGTGGGTGCCATTTATTGGCAAGA 251
Qy 277 GGGGGCCACCATCCGGAACATCAAAAACAGACCCAGTCCCAAGATAGACGTGCATAGGA 336
Db 252 GGGGGCCACCATCCGGAACATCAAAAACAGACCCAGTCCCAAGATAGACGTGCATAGGA 311
Qy 337 GAGAAACGAGGTGCAGCTGAAAAAGCCATCAGTGTGCACCTCCACCCCTGAGGGGTGCTC 396
Db 312 GAGAAACGAGGTGCAGCTGAAAAAGCCATCAGTGTGCACCTCCACCCCTGAGGGGTGCTC 371
Qy 397 CTCCGCTTGAAGATGATCTTCGAGATTATGCATTAAGAGGCTAAGGACACCAAAACCGC 456
Db 372 CTCCGCTTGAAGATGATCTTCGAGATTATGCATTAAGAGGCTAAGGACACCAAAACCGC 431
Qy 457 TGACGAGGTTCCCTGAGATCTCGGCCCATTAATCTTTGAGGGGCTCTCATTTGGCAA 516
Db 432 TGACGAGGTTCCCTGAGATCTCGGCCCATTAATCTTTGAGGGGCTCTCATTTGGCAA 491
Qy 517 GGAAGACCGAAACCTCGAAGAGGTAGAGCAAGATACCGAGACAAATAATCACCATCTCTCTC 576
Db 492 GGAAGACCGAAACCTCGAAGAGGTAGAGCAAGATACCGAGACAAATAATCACCATCTCTCTC 551
Qy 577 GTTGCAAGACCTTACCTTTTCAACCCCTGAGAGGACCATCATCTGTGAAGGGGCCATCGA 636
Db 552 GTTGCAAGACCTTACCTTTTCAACCCCTGAGAGGACCATCATCTGTGAAGGGGCCATCGA 611
Qy 637 GAATTG 642
Db 612 GAATTG 617

RESULT 4
B0648383
LOCUS AGENCOURT 8342265 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6268202
DEFINITION 5', mRNA sequence.
ACCESSION B0648383
VERSION B0648383.1 GI:21772555
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 950)
AUTHORS NIH-MGC http://mgs.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: CGAP (Stanford)
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LNCM2439 row: m column: 03
High quality sequence stop: 599.

FEATURES
source

Location/Qualifiers
1. 950
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clones="IMAGE:6268202"
/tissue type="hepatocellular carcinoma, cell line"
/lab host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_100"
/note="Organ: liver; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."

ORIGIN

Query Match 34.2%; Score 584; DB 13; Length 950;
Best Local Similarity 100.0%; Pred. No. 6.3e-290; Indels 0; Gaps 0;
Matches 584; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 37 AGCCATCATGAAGTGAATGCCACCAGTTGGAGAACCATGCGCTGAAGTCTCTTACAT 96
DB 12 AGCCATCATGAAGTGAATGCCACCAGTTGGAGAACCATGCGCTGAAGTCTCTTACAT 71

QY 97 CCCCAGTACAGATACACAGGACCTGAGATGGCGCGCGAGGGGCTTGGCTCTCG 156
DB 72 CCCCAGTACAGATACACAGGACCTGAGATGGCGCGCGAGGGGCTTGGCTCTCG 131

QY 157 GGGTCAGCCCCGCGAGGGCTTCACTGTGGCAGCGGGGGCCCCAGCCAGCAGCAAGT 216
DB 132 GGGTCAGCCCCGCGAGGGCTTCACTGTGGCAGCGGGGGCCCCAGCCAGCAGCAAGT 191

QY 217 GGACATCCCCCTTCGGCTCTGGTCCACCAGTATGGGTGGCCATTATGGCAAGGA 276
DB 192 GGACATCCCCCTTCGGCTCTGGTCCACCAGTATGGGTGGCCATTATGGCAAGGA 251

QY 277 GGGGGCCACCATCCGCAACATCAGACCCAGTCCCAAGATAGACGCTTGGCTCTCG 336
DB 252 GGGGGCCACCATCCGCAACATCAGACCCAGTCCCAAGATAGACGCTTGGCTCTCG 311

QY 337 GGAGAAAGCAGGTGCAGTGAAGGCTTCACTGTGGCAGCGGGGGCTTGGCTCTCG 396
DB 312 GGAGAAAGCAGGTGCAGTGAAGGCTTCACTGTGGCAGCGGGGGCTTGGCTCTCG 371

QY 397 CTCGGCTTGAAGATGATCTTGGAGATTCAGATAGAGGCTTAAGGACACCAACGGC 456
DB 372 CTCGGCTTGAAGATGATCTTGGAGATTCAGATAGAGGCTTAAGGACACCAACGGC 431

QY 457 TGACGAGGTTCCCTTGAAGATTCCTGGGCCATATTAATTTGTAGGGGCTCTCATGGCAA 516
DB 432 TGACGAGGTTCCCTTGAAGATTCCTGGGCCATATTAATTTGTAGGGGCTCTCATGGCAA 491

QY 517 GGRAGGACGGAACCTGAAGAGGTAGAGCAAGATACCGACAAATACCATCTCTTC 576
DB 492 GGRAGGACGGAACCTGAAGAGGTAGAGCAAGATACCGACAAATACCATCTCTTC 551

QY 577 GTTGCAAGACCTTACCTTTACACCTGAGAGGACCATCACTG 620
DB 552 GTTGCAAGACCTTACCTTTACACCTGAGAGGACCATCACTG 595

RESULT 5
BQ643920
LOCUS
DEFINITION
BQ643920
AGENCOURT_8303966 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6269413
5', mRNA sequence.
ACCESSION
BQ643920.1 GI:21768092
KEYWORDS
EST.

SOURCE
ORGANISM

Homo sapiens (human)

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 983)

AUTHORS

NIH-MGC <http://mgc.mci.nih.gov/>.

JOURNAL

Unpublished (1999)

COMMENT

Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: CGAP (Stanford)
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Plate: LLCN2442 row: 0 column: 14
High quality sequence stop: 491.

FEATURES

source

1. 983
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clones="IMAGE:6269413"
/tissue type="hepatocellular carcinoma, cell line"
/lab host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_100"
/note="Organ: liver; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."

ORIGIN

Query Match 31.9%; Score 545; DB 13; Length 983;
Best Local Similarity 99.8%; Pred. No. 9.3e-270;
Matches 595; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 37 AGCCATCATGAAGTGAATGCCACCAGTTGGAGAACCATGCGCTGAAGTCTCTTACAT 96
DB 12 AGCCATCATGAAGTGAATGCCACCAGTTGGAGAACCATGCGCTGAAGTCTCTTACAT 71

QY 97 CCCCAGTACAGATAGACAGGACCTGAGATGGCGCGCGAGGGGCTTGGCTCTCG 156
DB 72 CCCCAGTACAGATAGACAGGACCTGAGATGGCGCGCGAGGGGCTTGGCTCTCG 131

QY 157 GGGTCAGCCCCGCGAGGGCTTCACTGTGGCAGCGGGGGCCCCAGCCAGCAGCAAGT 216
DB 132 GGGTCAGCCCCGCGAGGGCTTCACTGTGGCAGCGGGGGCCCCAGCCAGCAGCAAGT 191

QY 217 GGACATCCCCCTTCGGCTCTGGTCCACCAGTATGGGTGGCCATTATGGCAAGGA 276
DB 192 GGACATCCCCCTTCGGCTCTGGTCCACCAGTATGGGTGGCCATTATGGCAAGGA 251

QY 277 GGGGGCCACCATCCGCAACATCAGACCCAGTCCCAAGATAGACGCTTGGCTCTCG 336
DB 252 GGGGGCCACCATCCGCAACATCAGACCCAGTCCCAAGATAGACGCTTGGCTCTCG 311

QY 337 GGAGAAAGCAGGTGCAGTGAAGGCTTCACTGTGGCAGCGGGGGCTTGGCTCTCG 396
DB 312 GGAGAAAGCAGGTGCAGTGAAGGCTTCACTGTGGCAGCGGGGGCTTGGCTCTCG 371

QY 397 CTCGGCTTGAAGATGATCTTGGAGATTCAGATAGAGGCTTAAGGACACCAACGGC 456
DB 372 CTCGGCTTGAAGATGATCTTGGAGATTCAGATAGAGGCTTAAGGACACCAACGGC 431

QY 457 TGACGAGGTTCCCTTGAAGATTCCTGGGCCATATTAATTTGTAGGGGCTCTCATGGCAA 516
DB 432 TGACGAGGTTCCCTTGAAGATTCCTGGGCCATATTAATTTGTAGGGGCTCTCATGGCAA 491

Db 432 TGACGAGGTTCCCTGAAGATCTCTGGCCCATATAAATCTTTGTAGGGCGTCTCATTTGGCAA 491

Qy 517 GGAAGGACGAACTGTAAGAGGTAGAGCAAGATACCGAGACAAATAATCACCATCTCTCTC 576

Db 492 GGAAGGACGAACTGTAAGAGGTAGAGCAAGATACCGAGACAAATAATCACCATCTCTCTC 551

Qy 577 GTTCCAAAGACCTTACCCCTTTACAAACCTTGAGAGGACCATCATCTGTGAAGGGGGCCA 632

Db 552 GTTCCAAAGACCTTACCCCTTTACAAACCTTGAGAGGACCATCATCTGTGAAGGGGGCCA 607

RESULT 6

LOCUS BGI15319

DEFINITION 602316274f1 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:4416354 5', mRNA sequence.

ACCESSION BGI15319

VERSION BGI15319.1 GI:12608925

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 831)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D. Email: cgapbs-remail.nih.gov Tissue Procurement: ATCC cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL) DNA Sequencing by: Incyte Genomics, Inc. Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov> Plate: LHM10147 row: d column: 19 High quality sequence stop: 731. Location/Qualifiers 1..831 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="IMAGE:4416354" /tissue_type="duodenal adenocarcinoma, cell line" /lab_host="DH10B (phage-resistant)" /clone_lib="NIH_MGC_88" /note="Organ: small intestine; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 1.767 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC Library."

FEATURES source

ORIGIN

Query Match 30.8%; Score 526; DB 12; Length 831;

Best Local Similarity 100.0%; Pred. No. 5.9e-260;

Matches 526; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 820 CTCCTTTATGACGCTCCGAGCAGGAGATGGTGCAGGTGTTATCCCCCGCCAGGAGT 879

Db 1 CTCCTTTATGACGCTCCGAGCAGGAGATGGTGCAGGTGTTATCCCCCGCCAGGAGT 60

Qy 880 GGGCGCATCATCGCAGAGAGGGGAGCAGCATCAACAGCTCTCCCGGTTCCAGCGC 939

Db 61 GGGCGCATCATCGCAGAGAGGGGAGCAGCATCAACAGCTCTCCCGGTTCCAGCGC 120

Qy 940 CTCATCAAGATTGACACCCGAAACACCTGACTCCAAAGTTTGTATGTTATCATCAC 999

Db 121 CTCATCAAGATTGACACCCGAAACACCTGACTCCAAAGTTTGTATGTTATCATCAC 180

Qy 1000 TGGACCGCCAGAGCCCAATTCAAGGCTCAGGGAGAGATCTATGGCAAACTCAAGAGGA 1059

Db 181 TGGACCGCCAGAGCCCAATTCAAGGCTCAGGGAGAGATCTATGGCAAACTCAAGAGGA 240

Qy 1060 GAACCTCTTGGTCCCAAGGAGGAGTGAAGCTGGAGACCCACATACCTGTGCCAGCATC 1119

Db 241 GAACCTCTTGGTCCCAAGGAGGAGTGAAGCTGGAGACCCACATACCTGTGCCAGCATC 300

Qy 1120 AGCAGCTGCGCGGGTCAATTTGGCAAAAGGTGGAAGAAACCGGTGAACGAGTTGCAGAAATTTGAC 1179

Db 301 AGCAGCTGCGCGGGTCAATTTGGCAAAAGGTGGAAGAAACCGGTGAACGAGTTGCAGAAATTTGAC 360

Qy 1180 GGCAGCTGAGGTGGTAGTACCAAGAGACAGACCCCTGATGAGACGACCCAGGTATCGT 1239

Db 361 GGCAGCTGAGGTGGTAGTACCAAGAGACAGACCCCTGATGAGACGACCCAGGTATCGT 420

Qy 1240 GAAATATCATCGGACATTTCTATGCCAGTCAAGTGGCTCAACGGAAGATCCGAGACATCCT 1299

Db 421 GAAATATCATCGGACATTTCTATGCCAGTCAAGTGGCTCAACGGAAGATCCGAGACATCCT 480

Qy 1300 GGGCCAGGTTAAGCGAGCATCATGAGGACAGAGTAACCAAGGCC 1345

Db 481 GGGCCAGGTTAAGCGAGCATCATGAGGACAGAGTAACCAAGGCC 526

RESULT 7

LOCUS BX327672/c

DEFINITION BX327672 Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens cDNA clone CS0DC018Y118 5-PRIME, mRNA sequence.

ACCESSION BX327672

VERSION BX327672.1 GI:30342696

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 919)

AUTHORS Li W.B., Gruber C., Jessee J. and Polayes D.

TITLE Full-length cDNA libraries and normalization

JOURNAL Unpublished (2001)

COMMENT Contact: Genoscope Genoscope - Centre National de Sequencage BP 191 91006 EVRY cedex - France Email: secref@genoscope.cns.fr, Web : www.genoscope.cns.fr Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 8684.r For more information about this cluster, see <http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS0BAG007ZH09> CS00648 1kcluster=8684.r. Contact : Feng Liang Email : fliang@lifetech.com URL : <http://fulllength.invitrogen.com/InvitroGen> Corporation 1600 Paraday Avenue Genoscope sequence ID : CS0BAG007ZH09_CS00648_1. Location/Qualifiers 1..919 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="CS0DC018Y118" /tissue_type="NEUROBLASTOMA COT 25-NORMALIZED" /note="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

FEATURES source

ORIGIN

Query Match 30.2%; Score 515; DB 13; Length 919;

Best Local Similarity 99.8%; Pred. No. 2.9e-254;

Matches 565; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 37 AGCCATCATGAGCTGAATGGCCACAGTTGAGAACCATGCGCTGGAAGGTCTCTACAT 96

Db 615 AGCCATCATGAGCTGAATGGCCACAGTTGAGAACCATGCGCTGGAAGGTCTCTACAT 556

Qy 97 CCCCAGTATGACATGACACAGGGACCTTGAGAAATGGGCGCCAGGGGGCTTTGGTCTCG 156

```

Db      555  CCCCAGATGAGCAGATAGCAGCAGGAGCTGAGATGGCGCGCGGAGGGGCTTTGGCTCTCG 496
Qy      157  GGGTCAGCCCCCGCAGGGCTCAGCTGTGGCAGCGGGGGCCCCCAGCCAGCAGCAGCAAGT 216
Db      495  GGGTCAGCCCCCGCAGGGCTCAGCTGTGGCAGCGGGGGCCCCCAGCCAGCAGCAGCAAGT 436
Qy      217  GGACATCCCTCTCGGCTCTGGTCCCGCCAGCCAGTATGTGGTGCCATTATTTGGCAAGGA 276
Db      435  GGACATCCCTCTCGGCTCTGGTCCCGCCAGCCAGTATGTGGTGCCATTATTTGGCAAGGA 376
Qy      277  GGGGGCCACCATCCGCAACATACAAAACAGACCCAGTCCAAAGATAGACGTGATAGGAA 336
Db      375  GGGGGCCACCATCCGCAACATACAAAACAGACCCAGTCCAAAGATAGACGTGATAGGAA 316
Qy      337  GGAGAAAGCAGGTGCAGCTGAAAAGCCATCAGTGTGCATCTCCAGCCCTGAGGGGTGCTC 396
Db      315  GAAGAACGCAAGTGCAGCTGAAAAGCCATCAGTGTGCATCTCCAGCCCTGAGGGGTGCTC 256
Qy      397  CTCGGCTTGAAGATGATCTTGGAGATTATGCATAAAGAGGCTTAAGACACCAAAACGGC 456
Db      255  CTCGGCTTGAAGATGATCTTGGAGATTATGCATAAAGAGGCTTAAGACACCAAAACGGC 196
Qy      457  TGACAGAGTTCCTCGAAGATCTCGGCCATAATACTTTGTAGGGGCTCTCATTTGGCAA 516
Db      195  TGACAGAGTTCCTCGAAGATCTCGGCCATAATACTTTGTAGGGGCTCTCATTTGGCAA 136
Qy      517  GGAAGGACGGAACCTGAGAGGTAGAGCAAGATACCGACACAAAATCACCATCTCTC 576
Db      135  GGAAGGACGGAACCTGAGAGGTAGAGCAAGATACCGACACAAAATCACCATCTCTC 76
Qy      577  GTTCCAGACCTTACCTTTACCAACC 602
Db      75  GTTCCAGACCTTACCTTTACCAACC 50

RESULT 8
LOCUS   BF984962
DEFINITION 602308484F1 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:4399556 5',
mRNA sequence.
ACCESSION BF984962
VERSION   BF984962.1 GI:12387774
KEYWORDS EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAMi0103 row: h column: 21
High quality sequence stop: 652.
Location/Qualifiers
1. .896
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4399556"
/tissue_type="duodenal adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_88"
/note="Organ: small intestine; Vector: pCMV-SPORT6;

FEATURES
source

```

```

Site_1: NotI; Site_2: SalI; Cloned unidirectionally;
oligo-dr primed. Average insert size 1.767 kb. Library
enriched for full-length clones and constructed by Life
Technologies. Note: this is a NIH_MGC Library."

Query Match      29 2%; Score 499; DB 10; Length 896;
Best Local Similarity 100.0%; Pred. No. 5.5e-246; Indels 0; Gaps 0;
Matches 499; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

ORIGIN
Qy      713  TGTCACTGATCCCTGGCTGAACTGCTGCTGTAGTCTTTTCCAGAGTTTATCCAGC 772
Db      137  TGTCACTGATCCCTGGCTGAACTGCTGCTGTAGTCTTTTCCAGAGTTTATCCAGC 196
Qy      773  GCAGTCCCGCGCTCCCGCAGCAGGTTACTGGGCTGCTCCCTATAGTCTCTTTATGCG 832
Db      197  GCAGTCCCGCGCTCCCGCAGCAGGTTACTGGGCTGCTCCCTATAGTCTCTTTATGCG 256
Qy      833  GCTCCGAGCAGGAGATGCTGAGGTGTTTATCCCGCCAGGAGTGGGCGCCATCATC 892
Db      257  GCTCCGAGCAGGAGATGCTGAGGTGTTTATCCCGCCAGGAGTGGGCGCCATCATC 316
Qy      893  GCGAAGAAAGGGCAGCAGCATCAAAACAGCTCTCCCGGTTTGGCAGCGCTCCATCAAGATT 952
Db      317  GCGAAGAAAGGGCAGCAGCATCAAAACAGCTCTCCCGGTTTGGCAGCGCTCCATCAAGATT 376
Qy      953  GCACCACCCGAAACACCTGACTCCAAAGTTTCTGTATGTTTATCATCTGGACCGCCAGAG 1012
Db      377  GCACCACCCGAAACACCTGACTCCAAAGTTTCTGTATGTTTATCATCTGGACCGCCAGAG 436
Qy      1013  GCCCAATTCAAGGCTCAGGGGAAGATCTATGCAAACTCAAGAGGAGAACTTTCTTTGGT 1072
Db      437  GCCCAATTCAAGGCTCAGGGGAAGATCTATGCAAACTCAAGAGGAGAACTTTCTTTGGT 496
Qy      1073  CCCAAGGAGGAAGTGAAGCTGGAGACCCACATAGTGTGCGAGCATCAGCAGCTGGCCGG 1132
Db      497  CCCAAGGAGGAAGTGAAGCTGGAGACCCACATAGTGTGCGAGCATCAGCAGCTGGCCGG 556
Qy      1133  GTCATTGTCAAAAGGTGGAACAAACGGTGAACAGTTCAGAGTTTTCAGCGGAGCTGAGGTG 1192
Db      557  GTCATTGTCAAAAGGTGGAACAAACGGTGAACAGTTCAGAGTTTTCAGCGGAGCTGAGGTG 616
Qy      1193  GTAGTACCAAGAGACCAGA 1211
Db      617  GTAGTACCAAGAGACCAGA 635

RESULT 9
LOCUS   BM561057
DEFINITION 1042 bp mRNA linear EST 20-FEB-2002
5', mRNA sequence.
ACCESSION BM561057
VERSION   BM561057.1 GI:18805998
KEYWORDS EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAMi2263 row: 1 column: 24

```

High quality sequence stop: 682.

FEATURES

Location/Qualifiers
1..1042
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5850887"
/tissue_type="retinoblastoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_67"
/note="Organ: eye; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo dr. Average insert size 1.75 kb. Library constructed by Life Technologies."

ORIGIN

Query Match 27.8%; Score 474; DB 12; Length 1042;
Best Local Similarity 100.0%; Pred. No. 4.8e-233;
Matches 474; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
713 TCTCAGCTGATCCCTGGCCTGAACCTGGCTGTAGGTCTTTTCCAGGCTTCATCCAGC 772
Db TCTCAGCTGATCCCTGGCCTGAACCTGGCTGTAGGTCTTTTCCAGGCTTCATCCAGC 342
773 GCAGTCCCGCGCTCCACAGAGGTTACTTGGGCTGCTCCCTATAGCTCCTTTATGCAG 832
Db GCAGTCCCGCGCTCCACAGAGGTTACTTGGGCTGCTCCCTATAGCTCCTTTATGCAG 402
833 GCTCCCGAGGAGGAGTGGTGCAGTGTATATCCCGCCAGCAGTGGCGGCATCATC 892
Db GCTCCCGAGGAGGAGTGGTGCAGTGTATATCCCGCCAGCAGTGGCGGCATCATC 462
893 GGCAGAAGGGGAGCAGCATCAACAGCTCTCCGGTTTCCAGCGCTCCATCAAGATT 952
Db GGCAGAAGGGGAGCAGCATCAACAGCTCTCCGGTTTCCAGCGCTCCATCAAGATT 522
953 GCACACCCGNAACACCTGACTCAAGTTCGTATGTTATCATCTGACACCGCCAGAG 1012
Db GCACACCCGNAACACCTGACTCAAGTTCGTATGTTATCATCTGACACCGCCAGAG 582
1013 GCCCAATTCAAGGCTCAGGGAAGAACTATGGCAAACTCAAGGAGGAGAACTCTTTGGT 1072
Db GCCCAATTCAAGGCTCAGGGAAGAACTATGGCAAACTCAAGGAGGAGAACTCTTTGGT 642
1073 CCCAAGGAGGAAGTGAAGCTGGAGACCCACATAGCTGTGCAGCATCAGAGCTGGCCGG 1132
Db CCCAAGGAGGAAGTGAAGCTGGAGACCCACATAGCTGTGCAGCATCAGAGCTGGCCGG 702
1133 GTCATTGGCAAAGTGGAAAACGGTGAACGAGTTGCAGAAATTTGACGGCAGCT 1186
Db GTCATTGGCAAAGTGGAAAACGGTGAACGAGTTGCAGAAATTTGACGGCAGCT 756

RESULT 10

BQ647561 1061 bp mRNA linear EST 15-JUL-2002
LOCUS AGENCOURT_8349614 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6284767
5', mRNA sequence.
ACCESSION BQ647561
VERSION BQ647561.1 GI:21771733
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE

1 (bases 1 to 1061)
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: CGAP (Stanford)
cDNA Library Preparation: Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCW2482 row: o column: 08
High quality sequence stop: 443.
Location/Qualifiers
1..1061
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6284767"
/tissue_type="hepatocellular carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_100"
/note="Organ: liver; Vector: pOTB7; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dr priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."

FEATURES

Source

ORIGIN

Query Match 24.9%; Score 426; DB 13; Length 1061;
Best Local Similarity 100.0%; Pred. No. 3.2e-208;
Matches 426; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
37 AGCCATCATGAAGCTGAATGGCCACAGTGTGGAGAACCATGCTCTGAGGTCTCTACAT 96
Db AGCCATCATGAAGCTGAATGGCCACAGTGTGGAGAACCATGCTCTGAGGTCTCTACAT 71
97 CCCCATGAGCAGATAGCAGGACCTGAGATCGCGCCGAGGGGCTTTGGCTCTCG 156
Db CCCCATGAGCAGATAGCAGGACCTGAGATCGCGCCGAGGGGCTTTGGCTCTCG 131
157 GGGTCAGCCCCCGCCAGGGCTCACCTGTGGCAGCGGGGGCCCCAGCCAGCAGCAAGT 216
Db GGGTCAGCCCCCGCCAGGGCTCACCTGTGGCAGCGGGGGCCCCAGCCAGCAGCAAGT 191
217 GCACATCCCCCTTCGGCTCTGGTCCACCCAGTATGTGGTGCCATTATGGCAAGA 276
Db GCACATCCCCCTTCGGCTCTGGTCCACCCAGTATGTGGTGCCATTATGGCAAGA 251
277 GGGGGCCACCATCCGCACATCACAAAACAGACCCAGTCCAGATAGAGTAGAGGAA 336
Db GGGGGCCACCATCCGCACATCACAAAACAGACCCAGTCCAGATAGAGTAGAGGAA 311
337 GGAGAACGAGGTGAGTGAAGGCAATCATGTGCACTCCACCCCTGAGGGGTGCTC 396
Db GGAGAACGAGGTGAGTGAAGGCAATCATGTGCACTCCACCCCTGAGGGGTGCTC 371
397 CTCGGCTTCTAAGATGATCTTCGAGATTATGATTAAGAGGCTAAAGGACCAAAAACGGC 456
Db CTCGGCTTCTAAGATGATCTTCGAGATTATGATTAAGAGGCTAAAGGACCAAAAACGGC 431
457 TCACGA 462
432 TGACGA 437

RESULT 11

BQ225582
LOCUS AGENCOURT_7593363 NIH_MGC_70 Homo sapiens cDNA clone IMAGE:6020837
DEFINITION 5', mRNA sequence.
ACCESSION BQ225582
VERSION BQ225582.1 GI:20406982
KEYWORDS EST.
SOURCE Homo sapiens (human)

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 712)
 AUTHORS NIH-MGC <http://mgs.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: L14M13225 row: j column: 06
 High quality sequence stop: 625.
 FEATURES
 source
 1..712
 Location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6020837"
 /tissue_type="epithelioid carcinoma"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH_MGC_70"
 /note="Organ: pancreas; Vector: pCMV-SPORT6; Site 1: Not I;
 Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 1.1 kb. Library constructed by Life
 Technologies."
 ORIGIN
 Query Match 24.5%; Score 419; DB 13; Length 712;
 Best Local Similarity 99.3%; Pred. No. 1.2e-204;
 Matches 669; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 QY 1031 GGAAGAATCTATGCAAACTCAAGGAGGAAGTCTTTGGTCCCAAGGAGGAAGTGAAG 1090
 Db 1 GGAAGAATCTATGCAAACTCAAGGAGGAAGTCTTTGGTCCCAAGGAGGAAGTGAAG 60
 QY 1091 CTGGAGACCCACATACCTGTGCGAGCATACGAGCTGGCCGGGTCTATGCGCAAGGTGA 1150
 Db 61 CTGGAGACCCACATACCTGTGCGAGCATACGAGCTGGCCGGGTCTATGCGCAAGGTGA 120
 QY 1151 AAAACGGTGAACGAGTTCGAGAAATTTGACCGCAGCTGAGGTGGTAGTACCAGAGACCAG 1210
 Db 121 AAAACGGTGAACGAGTTCGAGAAATTTGACCGCAGCTGAGGTGGTAGTACCAGAGACCAG 180
 QY 1211 ACCCTGATGAGACGACGAGTTCATCGTGAATAATCATCGGACATTTCTATGCCAGTCAG 1270
 Db 181 ACCCTGATGAGACGACGAGTTCATCGTGAATAATCATCGGACATTTCTATGCCAGTCAG 240
 QY 1271 ATGGCTCAACGGAGATCCGAGACATCCTGGCCAGGTTAAGCAGCAGCATCAGAGGGA 1330
 Db 241 ATGGCTCAACGGAGATCCGAGACATCCTGGCCAGGTTAAGCAGCAGCATCAGAGGGA 300
 QY 1331 CAGAGTAACGAGCCGACGAGGACGAGGAAGTGAACGACCCCTCCCTGTCCTTNGAGTCC 1390
 Db 301 CAGAGTAACGAGCCGACGAGGACGAGGAAGTGAACGACCCCTCCCTGTCCTTNGAGTCC 360
 QY 1391 AGGACAAACGCGGCAAAATCCAGAGTGTGCTCTCCCGCAGGCTCGAGATGAGTGG 1450
 Db 361 AGGACAAACGCGGCAAAATCCAGAGTGTGCTCTCCCGCAGGCTCGAGATGAGTGG 420
 QY 1451 GAATCCGGGACANWTGGCCGGGTGTAGATCAGGTGTGGCCACTTGATGAGAAAGATG 1510
 Db 421 GAATCCGGGACACTGGGGCCGGGTGTAGATCAGGTGTGGCCACTTGATGAGAAAGATG 480
 QY 1511 TTCCAGTGAGGAACCTTGATCTNTCAGCCCCCAACACCCACCAATGGCCCAACACTGT 1570
 Db 481 TTCCAGTGAGGAACCTTGATCTNTCAGCCCCCAACACCCACCAATGGCCCAACACTGT 540

QY 1571 NTGCCCTCGGGTGTCAAAATTTAGCCGCAAGGCACCTTTTAAACGTGGATTCTTTAAA 1630
 Db 541 CTGCCCTCGGGTGTCAAAATTTAGCCGCAAGGCACCTTTTAAACGTGGATTCTTTAAA 600
 QY 1631 GAAGCTCTCCAGGGCCCCCACCAGAGGGTGGATCACACCTCAGTGGGAGAAAAATAAAAT 1690
 Db 601 GAAGCTCTCCAGGGCCCCCACCAGAGGGTGGATCACACCTCAGTGGGAGAAAAATAAAAT 660
 QY 1691 TTCCTTCAGGTTTT 1704
 Db 661 TTCCTTCAGGTTTT 674
 RESULT 12
 LOCUS BQ647360
 DEFINITION AGENCOURT_8443628 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6284231
 5', mRNA sequence.
 ACCESSION BQ647360
 VERSION BQ647360
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 953)
 AUTHORS NIH-MGC <http://mgs.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: CGAP (Stanford)
 cDNA Library Preparation: Ruben Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: L14M2481 row: h column: 24
 High quality sequence stop: 651.
 FEATURES
 source
 1..953
 Location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6284231"
 /tissue_type="hepatocellular carcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH_MGC_100"
 /note="Organ: liver; Vector: pOTB7; Site 1: XhoI; Site 2:
 EcoRI; cDNA made by oligo-dT priming. Directionally cloned
 into EcoRI/XhoI sites using the following 5' adaptor:
 1.8kb. Library constructed by Ling Hong in the laboratory
 of Gerald M. Rubin (University of California, Berkeley)
 using ZAP-cDNA synthesis kit (Stratagene) and Superscript
 II RT (Life Technologies). Note: this is a NIH_MGC
 Library."
 ORIGIN
 Query Match 24.4%; Score 417; DB 13; Length 953;
 Best Local Similarity 100.0%; Pred. No. 1.4e-203;
 Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 37 AGCCATCATGAAGCTGAATGCCACCAAGTTGGAGAACCATGCCCTGAAGTCTCTCAT 96
 Db 305 AGCCATCATGAAGCTGAATGCCACCAAGTTGGAGAACCATGCCCTGAAGTCTCTCAT 364
 QY 97 CCCGATGAGCAGATAGCACAGGACCTGAGAAATGGCGCGCGGGGCTTTGGCTCTCG 156
 Db 365 CCCGATGAGCAGATAGCACAGGACCTGAGAAATGGCGCGCGGGGCTTTGGCTCTCG 424
 QY 157 GGGTCAGCCCCCGCAGGGCTCACCTGTGGCAGCGGGGGCCCCCAGCCAGCAGCAAGT 216

Db 425 GGGTCAGCCCGCCAGGGCTCACCTGTGGCAGCGGGGGCCCGAGCCNAGCAGCAAGT 484
Qy 217 GGACATCCCTTGGCTCTCTGGTGCCACCCAGATATGTGGTGCCATTTATGGCAAGA 276
Db 485 GGACATCCCTTGGCTCTCTGGTGCCACCCAGATATGTGGTGCCATTTATGGCAAGA 544
Qy 277 GGGGGCCACCATCGCAACATCAAAACAGAGCCAGTCCCAAGATAGAGTGCATAGGAA 336
Db 545 GGGGGCCACCATCGCAACATCAAAACAGAGCCAGTCCCAAGATAGAGTGCATAGGAA 604
Qy 337 GGAGAACGCGAGGTGAGCTGAAAGGACCATCAGTGTGCACTCCACCCCTGAGGGCTGCTC 396
Db 605 GGAGAACGCGAGGTGAGCTGAAAGGACCATCAGTGTGCACTCCACCCCTGAGGGCTGCTC 664
Qy 397 CTCGGCTTGAAGATGATCTTGAGATTATGCATAAAGAGGCTTAAGACACCAAAAC 453
Db 665 CTCGGCTTGAAGATGATCTTGAGATTATGCATAAAGAGGCTTAAGACACCAAAAC 721

RESULT 13
BU598549 884 bp mRNA linear EST 20-SEP-2002
LOCUS AGENCOURT_8822556 NIH_MGC_142 Homo sapiens cDNA clone IMAGE:6457432
DEFINITION 5', mRNA sequence.
ACCESSION BU598549
VERSION BU598549.1 GI:23250308
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 884)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: NCI
cDNA Library Preparation: Michael Brownstein Laboratory
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Plate: LSCM2639 row: e column: 17
High quality sequence stop: 420.
Location/Qualifiers
1..884
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6457432"
/tissue_types="mixed (pool of 40 RNAs)"
/lab_host="DH10B (TI-phage-resistant)"
/clone_lib="NIH_MGC_142"
/note="Vector: pDR-LIB; Site 1: SfiI (ggccattatggcc); Site 2: SfiI (ggccgctcgcc); Double-stranded cDNA was prepared from a pool of 40 cell line polyA+ RNAs (bladder - 2%, blood - 33.4%, brain - 5.6%, breast - 12.5%, colon - 4%, connective tissue - 1.4%, eye - 1%, intestine - 2.6%, kidney - 2.2%, liver - 5.7%, lung - 10.8%, NK-cell - 5.2%, ovary - 4%, pharynx - 2.5%, prostate - 4.3%, salivary gland - 1.3%, and skin - 2.3%). 5' and 3' adaptors were used in cloning as follows:
5'-AAGCAGTGGTATCAACGAGTGGCGATCGCATTCAGCGCGG-3' and 5'-ATTCTAGAGCGCGGCGGCGGCGGATG-dt(30)NN-3'. Full-length enriched library was constructed using the Clontech Creator SMART kit and size-selected to contain the >0.5 kb size fraction (other fractions present in NIH_MGC_141). Library created in the laboratory of M. Brownstein (NIH). Note: this is a NIH_MGC Library."

ORIGIN

Query Match 22.4%; Score 382; DB 13; Length 884;
Best Local Similarity 100.0%; Pred. No. 1.7e-185;
Matches 382; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 172 GGGCTCACTGTGGCAGCGGGGGCCCGAGCCNAGCAGCAAGT 231
Db 3 GGGCTCACTGTGGCAGCGGGGGCCCGAGCCNAGCAGCAAGT 62
Qy 232 GGTCTGTGTGCCACCCAGTATGTGGTGCCATTTATGGCAAGGAGGCGCCACCATCG 291
Db 63 GGTCTGTGTGCCACCCAGTATGTGGTGCCATTTATGGCAAGGAGGCGCCACCATCG 122
Qy 292 CAACATCAAAAACAGAGCCAGTCCCAAGATAGAGTGCATAGGAGGAGAGCGCAGGTGC 351
Db 123 CAACATCAAAAACAGAGCCAGTCCCAAGATAGAGTGCATAGGAGGAGAGCGCAGGTGC 182
Qy 352 AGCTGAAAAGCCATCAGTGTGCACTCCACCCCTGAGGGCTCTCTCCGCTTTGAAGAT 411
Db 183 AGCTGAAAAGCCATCAGTGTGCACTCCACCCCTGAGGGCTCTCTCCGCTTTGAAGAT 242
Qy 412 GATCTTGGAGATTATGCATAAAGAGGCTTAAGACACCAAAACGGCTGACGAGGTCCCT 471
Db 243 GATCTTGGAGATTATGCATAAAGAGGCTTAAGACACCAAAACGGCTGACGAGGTCCCT 302
Qy 472 GAAGATCTCTGGCCCATATAACTTTGTAGGGCGTCTCATTTGGCAAGGAGGACCGAACCT 531
Db 303 GAAGATCTCTGGCCCATATAACTTTGTAGGGCGTCTCATTTGGCAAGGAGGACCGAACCT 362
Qy 532 GAAGAAGGTAGAGCAAGATACC 553
Db 363 GAAGAAGGTAGAGCAAGATACC 384

RESULT 14
AW003366/c 761 bp mRNA linear EST 08-MAR-2000
LOCUS wq65h07.x1 NCI_CGAP_G66 Homo sapiens cDNA clone IMAGE:2476189 3'
DEFINITION similar to TR:O88477 O88477 CODING REGION DETERMINANT BINDING PROTEIN. ; mRNA sequence.
ACCESSION AW003366
VERSION AW003366.1 GI:5850282
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 761)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Ronaldo, Ph.D.
cDNA Sequencing by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html
Insert length: 1368 Std Error: 0.00
Seq primer: -400P from Gibco
High quality sequence stop: 456.
Location/Qualifiers
1..761
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2476189"
/tissue_type="pooled germ cell tumors"

```

/lab_host="DH10B"
/clone_lib="NCI_CGAP_GC6"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; Plasmid DNA
from the normalized library NCI_CGAP_GC4 was prepared, and
ss circles were made in vitro. Following HAP purification,
this DNA was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from a pool
of 5,000 clones made from the same library (clones IDs
1257096-1258531, 1469084-1470983, and 1475592-1476743).
Subtraction by Bento Soares and M. Fatima Bonaldo. "

ORIGIN
Query Match      17.3%; Score 295; DB 9; Length 761;
Best Local Similarity 100.0%; Pred. No. 1.5e-140;
Matches 295; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1024 GGCTCAGGGAAGAAATCTATGGCAAACTCAAGGAGGAGAACTTCTTTGGTCCCAAGGAGGA 1083
Db 611 GGCTCAGGGAAGAAATCTATGGCAAACTCAAGGAGGAGAACTTCTTTGGTCCCAAGGAGGA 552

Qy 1084 AGTGAAGCTGGAGACCCACATACGTGTGCCAGATCAGCAGCTGGCGGGTCAATTGGCAA 1143
Db 551 AGTGAAGCTGGAGACCCACATACGTGTGCCAGATCAGCAGCTGGCGGGTCAATTGGCAA 492

Qy 1144 AGGTGGAAAAACGGTGAACGAGTTGCAGATTTTACGGCAGCTGAGTGGTACCAAG 1203
Db 491 AGGTGGAAAAACGGTGAACGAGTTGCAGATTTTACGGCAGCTGAGTGGTACCAAG 432

Qy 1204 AGACCAACCCCTGATCAGAACCCAGCAGTCATCGTGAATCATCGGACATTTCTATGC 1263
Db 431 AGACCAACCCCTGATCAGAACCCAGCAGTCATCGTGAATCATCGGACATTTCTATGC 372

Qy 1264 CAGTCAGATGGCTCAACGGGAAGATCCGAGACATCCTGGCCCGAGTTAAGCAGCAG 1318
Db 371 CAGTCAGATGGCTCAACGGGAAGATCCGAGACATCCTGGCCCGAGTTAAGCAGCAG 317

RESULT 15
CB146278      568 bp      mRNA      linear      EST 29-JAN-2003
LOCUS        K-EST0201433 L16HLK3 Homo sapiens cDNA clone L16HLK3-20-H04 5',
DEFINITION   mRNA sequence.
ACCESSION   CB146278
VERSION      CB146278.1 GI:28126979
KEYWORDS     EST.
SOURCE       Homo sapiens (human)
ORGANISM     Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 568)
Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
Kim,Y.S.
21C Frontier Korean EST Project 2001
Unpublished (2002)
Contact: Kim YS
Genome Research Center
Korea Research Institute of Bioscience & Biotechnology
52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4409
Email: yongsung@mail.kribb.re.kr
Plate: 20 row: H column: 04
High quality sequence stop: 568.
Location/Qualifiers
1. 568
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="L16HLK3-20-H04"
/cell_line="HLK-3"
/lab_host="Top10F"
FEATURES
source

```

```

/clone_lib="L16HLK3"
/note="Organ: Liver; Vector: pT73D-Pac; Site 1: EcoRI;
Site 2: NotI; The library was contributed by the Soares
laboratory and it was constructed as described by Bonaldo,
M.F., Lennon, G. and Soares, M.B. (1996), Genome Research
6(9): 791-806. RNA was prepared from harvested cell
culture."

ORIGIN
Query Match      17.0%; Score 291; DB 14; Length 568;
Best Local Similarity 100.0%; Pred. No. 1.6e-138;
Matches 291; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 713 TCTCACCTGATCCCTGGCCTGAACCTGGCTGCTAGGTCTTTTCCAGGCTTCATCCAGC 772
Db 278 TCTCACCTGATCCCTGGCCTGAACCTGGCTGCTAGGTCTTTTCCAGGCTTCATCCAGC 337

Qy 773 GCAGTCCCGCGCCTCCAGAGCGTTACTGGGGCTGCTCCCTATAGTCTCTTTATGCAG 832
Db 338 GCAGTCCCGCGCCTCCAGAGCGTTACTGGGGCTGCTCCCTATAGTCTCTTTATGCAG 397

Qy 833 GCTCCCGAGCAGGAGATGCTGCAAGTGTATCCCGCCAGGCGAGTGGCGCCATCATC 892
Db 398 GCTCCCGAGCAGGAGATGCTGCAAGTGTATCCCGCCAGGCGAGTGGCGCCATCATC 457

Qy 893 GGCAGAGAGGGCGAGCAGCATCAAAAGCTCTCCCGTTTGGCAGCGCTCCATCAAGATT 952
Db 458 GGCAGAGAGGGCGAGCAGCATCAAAAGCTCTCCCGTTTGGCAGCGCTCCATCAAGATT 517

Qy 953 GCACACCCGAAACACCTGACTCCAAAGTTGTTATGTTATCATCTCACTGGA 1003
Db 518 GCACACCCGAAACACCTGACTCCAAAGTTGTTATGTTATCATCTCACTGGA 568

Search completed: July 14, 2004, 03:02:41
Job time : 4594 secs

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_r2p model

Run on: July 13, 2004, 12:24:58 ; Search time 30 Seconds
(without alignments)
10953.008 Million cell updates/sec

Title: US-09-270-437D-5
Perfect score: 560
Sequence: 1 agggagcgcgcgcacccgc.....attcttcagggttttaaaa 1708

Scoring table: OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 283366 seqs, 96191526 residues

Word size: 1

Total number of hits satisfying chosen parameters: 565818

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+r2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US09270437/tunat_13072004_121956_9997/app_query.fasta_1.1563
-DB=PIR_78 -QMT=fastan -SUFFIX=ipr -MINMATCH=0.1 -LOOPCPI=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc
-NORM=ext -HEAPSIZES=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09270437 @CGN_1_1_44 @tunat_13072004_121956_9997 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : PIR_78.*

1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	8	1.4	95	2 H95092	hypothetical prote
C 2	8	1.4	95	2 E97960	hypothetical prote
C 3	8	1.4	185	2 I39736	hypothetical prote
C 4	8	1.4	217	2 AD2683	Conserved hypotet
C 5	8	1.4	221	2 AF3195	transcription regu
C 6	8	1.4	223	2 A83859	menaquinol-cytochr
C 7	8	1.4	224	2 I39943	menaquinol-cytochr
C 8	8	1.4	227	2 A70036	capsular polysacch
C 9	8	1.4	247	2 B90733	probable major tai
C 10	8	1.4	249	2 D85583	probable tail comp
C 11	8	1.4	253	2 B97465	hypothetical prote
C 12	8	1.4	259	2 G64831	probable membrane
C 13	8	1.4	259	2 C90754	hypothetical prote
C 14	8	1.4	259	2 A85618	hypothetical prote

15	8	1.4	301	2 A42223	heterodisulfide re
16	8	1.4	306	2 A50876	agmatine ureohydro
17	8	1.4	340	2 AB2584	hypothetical prote
18	8	1.4	340	2 B97366	frCB protein (AF19
19	8	1.4	351	2 AC3090	conserved hypotet
20	8	1.4	351	2 G98196	hypothetical prote
21	8	1.4	357	2 B81396	probable aminotran
22	8	1.4	367	2 T36278	hypothetical prote
23	8	1.4	374	2 T07887	G box-binding prot
24	8	1.4	376	2 T07882	G box-binding prot
25	8	1.4	457	2 B87269	hypothetical prote
26	8	1.4	475	2 G70861	hypothetical prote
27	8	1.4	532	2 E95949	probable oligopept
28	8	1.4	546	2 A10278	probable AMP nucle
29	8	1.4	555	2 T30349	structural protein
30	8	1.4	573	2 C85433	splicing factor-1i
31	8	1.4	573	2 C86266	F3F19.21 protein -
32	8	1.4	600	2 T09676	probable ATP-depen
33	8	1.4	638	2 F75547	anthranilate synth
C 34	8	1.4	645	2 H96011	asparagine synthas
35	8	1.4	646	1 S15901	chromogranin B pre
36	8	1.4	680	2 T25832	hypothetical prote
C 37	8	1.4	735	2 A84829	hypothetical prote
C 38	8	1.4	755	2 T47731	probable integral
C 39	8	1.4	759	2 T39090	hypothetical prote
40	8	1.4	812	2 T34180	hypothetical prote
C 41	8	1.4	963	2 T26022	probable helicase
42	8	1.4	1030	2 S73460	protein C08H9.2 [i
43	8	1.4	1198	2 B88279	hypothetical prote
44	8	1.4	1220	2 T19117	hypothetical prote
45	8	1.4	1269	2 F88101	protein W09G10.4 [

ALIGNMENTS

RESULT 1

H95092 Hypothetical protein SP0800 [imported] - Streptococcus pneumoniae (strain TIGR4)
C:Species: Streptococcus pneumoniae
C>Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001
C:Accession: H95092

R:Tetelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Hei
on, J.D.; Unayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple,
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001

A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison
A>Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: H95092

A>Status: Preliminary
A:Molecule type: DNA

A:Residues: 1-95 <XUR>
A:Cross-references: GB:AB005672; PIDN:AAK74937.1; PID:G14972276; GSPDB:GN00164; TIGR:SP

A:Experimental source: strain TIGR4
C:Genetics:
A:Gene: SP0800

Alignment Scores:				
Pred. No.:	63.8	Length:	95	
Score:	8.00	Matches:	8	
Percent Similarity:	100.00%	Conservative:	0	
Best Local Similarity:	100.00%	Mismatches:	0	
Query Match:	1.44%	Indels:	0	
DB:	2	Gaps:	0	

US-09-270-437D-5 (1-1708) x H95092 (1-95)

QY 1702 AACCTGAAGGAATTTTATTTTC 1679

|||||
36 AsnLeuLysGluIleLeuPhephe 43

RESULT 2

E97960

hypothetical protein spr0709 [imported] - Streptococcus pneumoniae (strain R6)
C/Species: Streptococcus pneumoniae
C/Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 22-Oct-2001
C/Accession: E97960
R/Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; E
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M
y, P.; Sun, P.M.; Winkler, M.F.
J. Bacteriol. 183, 5709-5717, 2001
A/Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
A/Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A/Reference number: A97872; MUID:21429245; PMID:11544234
A/Accession: E97960
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-95 <KUR>
A/Cross-references: GB:AE007317; PIDN:AAK99513.1; PID:g15459299; GSPDB:GN00174
C/Genetics:
A/Gene: spr0709

Alignment Scores:
Pred. No.: 63.8 Length: 95
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.44% Indels: 0
DB: 2 Gaps: 0

US-09-270-437D-5 (1-1708) x E97960 (1-95)
QY 1702 AACCTGAAGGAAATTTATTTTC 1679
Db 36 AsnLeuLysgluileuLeuPhePhe 43
|||||

RESULT 3
I39736
hypothetical protein 6 - Anabaena variabilis
C/Species: Anabaena variabilis
C/Date: 19-Jul-1996 #sequence_revision 19-Jul-1996 #text_change 15-Oct-1999
C/Accession: S68185; I39736
R/Schmitz, O.; Boison, G.; Hilscher, R.; Hundeshagen, B.; Zimmer, W.; Lottspeich, F.; Bo
Eur. J. Biochem. 233, 266-276, 1995
A/Title: Molecular biological analysis of a bidirectional hydrogenase from cyanobacteria
A/Reference number: I39730; MUID:96061958; PMID:7588754
A/Accession: S68185
A/Status: nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-185 <SCH>
A/Cross-references: EMBL:X79285; MID:g1032475; PIDN:CAA55879.1; PID:g1032482
A/Experimental source: ATCC 29413
A/Note: the nucleotide sequence was submitted to the EMBL Data Library, May 1994

Alignment Scores:
Pred. No.: 58 Length: 185
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.44% Indels: 0
DB: 2 Gaps: 0

US-09-270-437D-5 (1-1708) x I39736 (1-185)
QY 478 GATCTTCAGGGGAACTCGTCAGC 455
Db 90 AspLeuGlnGlyAsnLeuValSer 97
|||||

RESULT 4
AD2683
Conserved hypothetical protein Atu0868 [imported] - Agrobacterium tumefaciens (strain C58)
C/Species: Agrobacterium tumefaciens
C/Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C/Accession: AD2683
R/Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClell

; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A/Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A/Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A/Reference number: AB2577; MUID:21608550; PMID:11743193
A/Accession: AD2683
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-217 <KUR>
A/Cross-references: GB:AE008688; PIDN:AAL41882.1; PID:g17739245; GSPDB:GN00186
A/Experimental source: strain C58 (Dupont)
C/Genetics:
A/Gene: Atu0868
A/Map position: circular chromosome

Alignment Scores:
Pred. No.: 56.7 Length: 217
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.44% Indels: 0
DB: 2 Gaps: 0

US-09-270-437D-5 (1-1708) x AD2683 (1-217)
QY 1661 TCCACCTCTTGTGGGGCTGGA 1638
Db 103 SerThrLeuLeuValGlyProGly 110
|||||

RESULT 5
AF3195
transcription regulator, TetR family Atu5291 [imported] - Agrobacterium tumefaciens (str
C/Species: Agrobacterium tumefaciens
C/Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C/Accession: AF3195
R/Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClell
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A/Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A/Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A/Reference number: AB2577; MUID:21608550; PMID:11743193
A/Accession: AF3195
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-221 <KUR>
A/Cross-references: GB:AE008687; PIDN:AAL45980.1; PID:g17743733; GSPDB:GN00188
A/Experimental source: strain C58 (Dupont)
C/Genetics:
A/Gene: Atu5291
A/Genome: plasmid

Alignment Scores:
Pred. No.: 56.6 Length: 221
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.44% Indels: 0
DB: 2 Gaps: 0

US-09-270-437D-5 (1-1708) x AF3195 (1-221)
QY 227 GGGGATGTCACATGCTGCTGCT 204
Db 111 GlyGlyCysProLeuAlaAlaAla 118
|||||

RESULT 6
AB3859
menaquinol-cytochrome c reductase (cytochrome b subunit) qcrB [imported] - Bacillus halo
C/Species: Bacillus halodurans
C/Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 03-Jun-2002

C/Accession: A83859
 R/Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira
 Nucleic Acids Res. 28, 4317-4331, 2000
 A/Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and
 A/Reference number: A83650; MUID:20512582; PMID:11058132
 A/Accession: A83859
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-223 <STO>
 A/Cross-references: GB:AP001512; GB:BA000004; NID:g10174030; PIDN:BA05392.1; GSPDB:GN00
 A/Experimental source: strain C-125
 C/Genetics:
 A/Gene: qcrB

Alignment Scores:
 Pred. No.: 56.5 Length: 223
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 1.43% Indels: 0
 DB: 2 Gaps: 0

US-09-270-437D-5 (1-1708) x A83859 (1-223)

QY 786 CTCCAGCGGTACTGGGGCTG 809
 Db 199 LeuProAlaAlaLeuGlyLeu 206

RESULT 7
 I39943
 menaquinol-cytochrome-c reductase (EC 1.10.2.-) cytochrome b6 qcrB - *Bacillus stearother*
 C/Species: *Bacillus stearothermophilus*
 C/Date: 19-Jul-1996 #sequence_revision 19-Jul-1996 #text_change 04-Mar-2000
 C/Accession: I39943
 R/Sone, N.; Sawa, G.; Sone, T.; Noguchi, S.
 J. Biol. Chem. 2, 10612-10617, 1995
 A/Title: Thermophilic bacilli have split cytochrome b genes for cytochrome b6 and subun
 A/Reference number: I39943
 A/Accession: I39943
 A/Status: preliminary; translated from GB/EMBL/DDBJ
 A/Molecule type: DNA
 A/Residues: 1-224 <RES>
 A/Cross-references: GB:D45410; NID:G902924; PID:G643677
 C/Genetics:
 A/Gene: petB
 A/Start codon: GNG
 C/Superfamily: cytochrome b6; cytochrome b6 homology
 C/Keywords: chromoprotein; electron transfer; heme; iron; metalloprotein; oxidoreductase
 F/16-224/Domain: cytochrome b6 homology <CB6>
 F/43/Binding site: heme (Cys) (covalent) (probably high potential) #status predicted
 F/94,196/Binding site: heme iron (His) (axial ligands) (low potential) #status predicted
 F/108,211/Binding site: heme iron (His) (axial ligands) (high potential) #status predicted

Alignment Scores:
 Pred. No.: 56.5 Length: 224
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 1.43% Indels: 0
 DB: 2 Gaps: 0

US-09-270-437D-5 (1-1708) x I39943 (1-224)

QY 786 CTCCAGCGGTACTGGGGCTG 809
 Db 200 LeuProAlaAlaLeuGlyLeu 207

RESULT 8
 A70036
 capsular polysaccharide biosynthesis homolog yveL - *Bacillus subtilis*
 C/Species: *Bacillus subtilis*
 C/Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 20-Jun-2000
 C/Accession: A70036

R/Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bette
 C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch
 A.; Ehrlich, S.D.; Emerson, P.T.; Enian, K.D.; Errington, J.; Fabret, C.; Ferrari, S.; Ch
 Nature 390, 249-256, 1997
 A/Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galle
 iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F
 Koetter, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois
 A/Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mauee
 Y. M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetell
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon
 A/Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Sero
 akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama
 T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, I
 A/Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*
 A/Reference number: A69580; MUID:98044033; PMID:9384377

A/Accession: A70036
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown
 A/Molecule type: DNA
 A/Residues: 1-227 <KUN>
 A/Cross-references: GB:Z99121; GB:AL009126; NID:g2635827; PIDN:CAB15441.1; PID:g2635949
 A/Experimental source: strain 168
 C/Genetics:
 A/Gene: yveL
 C/Superfamily: capsular polysaccharide biosynthesis protein cpsC

Alignment Scores:
 Pred. No.: 56.4 Length: 227
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 1.44% Indels: 0
 DB: 2 Gaps: 0

US-09-270-437D-5 (1-1708) x A70036 (1-227)

QY 1367 CTGTCATCTCTCCGTCGCTGGG 1344

Db 48 LeuValThrSerValProGly 55

RESULT 9

B90733

probable major tail protein [imported] - *Escherichia coli* (strain O157:H7, substrain R1
 C/Species: *Escherichia coli*
 C/Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
 C/Accession: B90733
 R/Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
 gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
 DNA Res. 8, 11-22, 2001
 A/Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and gen
 A/Reference number: A99629; MUID:21156231; PMID:11258796

A/Accession: B90733

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-247 <HAY>

A/Cross-references: GB:BA000007; PIDN:BA034257.1; PID:g13360293; GSPDB:GN00154

A/Experimental source: strain O157:H7, substrain RMD 050952

C/Genetics:

A/Gene: ECS0834

C/Superfamily: phage lambda major tail protein V

Alignment Scores:

Pred. No.: 55.7 Length: 247
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 1.43% Indels: 0
 DB: 2 Gaps: 0

US-09-270-437D-5 (1-1708) x B90733 (1-247)

QY 359 AAAGCCATCAGTGTGCACTCCACC 382

|||||

Db 193 LysAlaIleSerValHisSerThr 200
 RESULT 10
 D85583
 Probable tail component of prophage CP-933K Z0972 [imported] - Escherichia coli (strain C58)
 C:Species: Escherichia coli
 C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
 C:Accession: D85583
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
 Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
 Nature 409, 529-533, 2001
 A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
 A:Reference number: A85480; MUID:21074935; PMID:11206551
 A:Accession: D85583
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-249 <STO>
 A:Cross-references: GB:AE005174; NID:g12513740; PIDN:AA055128.1; GSPDB:GN00145; UWGP:Z09
 A:Experimental source: strain O157:H7, substrain EDL933
 C:Genetics:
 A:Gene: Z0972
 C:Superfamily: phage lambda major tail protein V

Alignment Scores:
 Pred. No.: 55.6 Length: 249
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 1.43% Indels: 0
 DB: 2 Gaps: 0

US-09-270-437D-5 (1-1708) x D85583 (1-249)
 Qy 359 AAAGCCATCAGTGTGCATCCACC 382
 Db 195 LysAlaIleSerValHisSerThr 202
 |||||

RESULT 11
 B97465
 Hypothetical protein AGR_C_1585 [imported] - Agrobacterium tumefaciens (strain C58, Cere
 C:Species: Agrobacterium tumefaciens
 C>Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 18-Nov-2002
 C:Accession: B97465
 R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorollo, B.; Goldman,
 A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
 Science 294, 2323-2328, 2001
 A:Title: Genome sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
 A:Reference number: A97359; MUID:21608551; PMID:11743194
 A:Accession: B97465
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-253 <KUR>
 A:Cross-references: GB:AE007869; PIDN:AAK86675.1; PID:g15155859; GSPDB:GN00169
 C:Genetics:
 A:Gene: AGR_C_1585
 A:Map position: circular chromosome

Alignment Scores:
 Pred. No.: 55.5 Length: 253
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 1.44% Indels: 0
 DB: 2 Gaps: 0

US-09-270-437D-5 (1-1708) x B97465 (1-253)
 Qy 1661 TCACCCCTCTTGGTGGGCTCGA 1638
 Db 139 SerThrLeuValGlyProGly 146
 |||||

RESULT 12
 G64831
 Probable membrane protein ybcC - Escherichia coli (strain K-12)
 C:Species: Escherichia coli
 C>Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002
 C:Accession: G64831
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
 A.; Rose, D.J.; Mau, B.; Shao, Y.
 Science 277, 1453-1462, 1997
 A:Title: The complete genome sequence of Escherichia coli K-12.
 A:Reference number: A64720; MUID:97426617; PMID:9278503
 A:Accession: G64831
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-259 <BLAT>
 A:Cross-references: GB:AE000194; GB:U00096; NID:g1787148; PIDN:AACT4006.1; PID:g1787150;
 A:Experimental source: strain K-12, substrain MG1655
 C:Genetics:
 A:Gene: ybcC
 C:Superfamily: conserved hypothetical protein aq_1986
 C:Keywords: transmembrane protein
 F:13-29/Domain: transmembrane #status predicted <TM1>
 F:39-55/Domain: transmembrane #status predicted <TM2>

Alignment Scores:
 Pred. No.: 55.3 Length: 259
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 1.44% Indels: 0
 DB: 2 Gaps: 0

US-09-270-437D-5 (1-1708) x G64831 (1-259)
 Qy 479 GGATCTTCAGGGGAACCTCGTCAG 456
 Db 251 GlySerSerGlyGluProArgGin 258
 |||||

RESULT 13
 C90754
 Hypothetical protein ECs1003 [imported] - Escherichia coli (strain O157:H7, substrain RI
 C:Species: Escherichia coli
 C>Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
 C:Accession: C90754
 R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
 gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
 A. Res. 8, 11-22, 2001
 A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno
 A:Reference number: A99629; MUID:21156231; PMID:11258796
 A:Accession: C90754
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-259 <HAY>
 A:Cross-references: GB:BA000007; PIDN:BA34426.1; PID:g13360462; GSPDB:GN00154
 A:Experimental source: strain O157:H7, substrain RIMD 0509952
 C:Genetics:
 A:Gene: ECs1003
 C:Superfamily: conserved hypothetical protein aq_1986

Alignment Scores:
 Pred. No.: 55.3 Length: 259
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 1.44% Indels: 0
 DB: 2 Gaps: 0

US-09-270-437D-5 (1-1708) x C90754 (1-259)
 Qy 479 GGATCTTCAGGGGAACCTCGTCAG 456
 Db 251 GlySerSerGlyGluProArgGin 258
 |||||

RESULT 14
 A85618

hypothetical protein ybcC [imported] - Escherichia coli (strain O157:H7, substrain EDL93)
 C/Species: Escherichia coli
 C/Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
 C/Accession: A85618
 R/Perna N.T.: Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
 Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimlant, E.; Potamousis, K.; Apodaca,
 Nature 409, 529-533, 2001
 A/Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
 A/Reference number: A85480; MUID:21074935; PMID:11206551
 A/Accession: A85618
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-259 <STO>
 A/Cross-references: GB:AE005174; NID:gl2514089; PIDN:AA055405.1; GSPDB:GN00145; UMG:P212
 A/Experimental source: strain O157:H7, substrain EDL933
 C/Genetics:
 A/Gene: ybcC
 C/Superfamily: conserved hypothetical protein aq_1986

Alignment Scores:
 Pred. No.: 55.3 Length: 259
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 1.44% Indels: 0
 DB: 2 Gaps: 0

US-09-270-437D-5 (1-1708) x A85618 (1-259)

QY 479 GGATCTTCAGGGGAACCTCGTCAG 456

Db 251 GlySerSerGlyGluProArgGln 258

RESULT 15

AF2223
 heterodisulfide reductase, chain B [imported] - Nostoc sp. (strain PCC 7120)
 C/Species: Nostoc sp. PCC 7120
 A/Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
 C/Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
 C/Accession: AF2223
 R/Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi
 Nakazaki, N.; Shimo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S
 DNA Res. 8, 205-213, 2001
 A/Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
 A/Reference number: AB1807; MUID:21595285; PMID:11759840
 A/Accession: AF2223
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-301 <XUR>
 A/Cross-references: GB:BA000019; PIDN:BA075040.1; PID:gl7132436; GSPDB:GN00179
 A/Experimental source: strain PCC 7120
 C/Genetics:
 A/Gene: all3341
 C/Superfamily: Acidianus ambivalens succinate dehydrogenase chain C

Alignment Scores:
 Pred. No.: 54.2 Length: 301
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 1.43% Indels: 0
 DB: 2 Gaps: 0

US-09-270-437D-5 (1-1708) x AF2223 (1-301)

QY 144 GCTTGGCTCTCGGGTCAGCC 167

Db 273 AlaleuAlaleuGlyValSerPro 280

Search completed: July 13, 2004, 12:35:45
 Job time : 37 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 13, 2004, 21:42:59 ; Search time 6722 Seconds
(without alignments)
11013.077 Million cell updates/sec

Title: US-09-270-437D-5
Perfect score: 1708
Sequence: 1 aggaagctgcgcacccgc.....atttccttcaggttttaaaa 1708

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0
Searched: 3470272 seqs, 21671516995 residues

Word size : 0
Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : GenEmbl.*

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.em.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.mu.*
- 20: em.om.*
- 21: em.or.*
- 22: em.ov.*
- 23: em.pat.*
- 24: em.ph.*
- 25: em.pl.*
- 26: em.ro.*
- 27: em.sts.*
- 28: em.un.*
- 29: em.vi.*
- 30: em.htg.hum.*
- 31: em.htg.in.v.*
- 32: em.htg.other.*
- 33: em.htg.mus.*
- 34: em.htg.pln.*
- 35: em.htg.rod.*
- 36: em.htg.mam.*
- 37: em.htg.vrt.*
- 38: em.sy.*
- 39: em.htgo.hum.*
- 40: em.htgo.mus.*
- 41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1708	100.0	1708	6	ARI171864 Sequence
2	1708	100.0	1708	6	ARI171864 Sequence
3	1708	100.0	1708	6	BD209924 Isolated
4	1672	97.9	1946	6	ARI171866 Sequence
5	1672	97.9	1946	6	ARI171866 Sequence
6	1672	97.9	1946	6	BD209926 Isolated
7	997	58.4	2130	9	AF117106 Homo sapi
8	821	48.1	2381	9	AF198254 Homo sapi
9	741	43.4	2780	6	BD127811 Primer fo
10	741	43.4	2780	9	AK074915 Homo sapi
11	414	24.2	833	6	BD124998 Primer fo
12	414	24.2	833	6	BD126884 Primer fo
13	284	16.6	107848	9	AC105030 Homo sapi
14	284	16.6	154604	2	AC104974 Homo sapi
15	284	16.6	159122	2	AC025556 Homo sapi
16	284	16.6	168613	9	AC091133 Homo sapi
17	208	12.2	65435	2	AC104587 Homo sapi
18	186	10.9	65435	2	AC104587 Homo sapi
19	136	8.0	164950	2	AC105191 Homo sapi
20	121	7.1	67462	2	AC091595 Homo sapi
21	62	3.6	67462	2	AC091595 Homo sapi
22	38	2.2	216616	10	AC094527 Rattus no
23	38	2.2	220434	2	AC106496 Rattus no
24	38	2.2	232468	2	AC094487 Rattus no
25	36	2.1	662	9	HS331847 Homo sapi
26	36	2.1	845	9	HS334465 Homo sapi
27	32	1.9	2105	10	AF541940 Rattus no
28	32	1.9	196392	2	AC120322 Rattus no
29	31	1.8	2223	10	AF061569 Mus muscu
30	31	1.8	2224	6	ARI60244 Sequence
31	31	1.8	2444	10	EC051679 Mus muscu
32	31	1.8	71109	10	AL606704 Mouse DNA
33	31	1.8	188834	10	AC098642 Genomic s
34	31	1.8	239837	10	AC084407 Mus Muscu
35	26	1.5	187119	10	AC123533 Mus muscu
36	26	1.5	219696	2	AC099577 Mus muscu
37	25	1.5	177384	2	AC091786 Mus muscu
38	24	1.4	24	6	ARI60252 Sequence
39	23	1.3	1493	23	ARI60250 Sequence
40	23	1.3	178165	2	AF076930 Mus muscu
41	23	1.3	178165	2	AC109259 Mus muscu
42	23	1.3	180104	9	AF123462 Homo sapi
43	22	1.3	1536	6	E00532 DNA encodin
44	22	1.3	1787	8	X01104 Soybean hea
45	22	1.3	170348	2	AC105168 Mus muscu

ALIGNMENTS

RESULT 1	ARI171864	Sequence 5 from patent US 6297364.	1708 bp	DNA	linear	PAT 17-DEC-2001
LOCUS	ARI171864	Sequence 5 from patent US 6297364.				
DEFINITION	ARI171864					
ACCESSION	ARI171864					
VERSION	ARI171864.1	GI:17910814				
KEYWORDS						
SOURCE	Unknown.					
ORGANISM	Unknown.					
REFERENCE	1 (bases 1 to 1708)					
AUTHORS	Chen,Y.-T., Gure,A., Tsang,S., Stockert,E., Jager,E., Alexander,K. and Old,L.J.					
TITLE	Isolated nucleic acid molecule encoding cancer associated antigen, the antigen itself, and uses thereof					

JOURNAL Patent: US 6297364-A 5 02-OCT-2001;
FEATURES Location/Qualifiers
source 1..1708
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN

Query Match 100.0%; Score 1708; DB 6; Length 1708;
Best local similarity 100.0%; Pred. No. 0;
Matches 1708; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGAGCGTCCCGACCGCCAGTTTACCCGGGGAGCCATCATGAAGCTGAATGGCCA 60
DB 1 AGGAGCGTCCCGACCGCCAGTTTACCCGGGGAGCCATCATGAAGCTGAATGGCCA 60

QY 61 CCAGTTGGAGAACCATGCCCTGAAGTCTCTACATCCCGATGAGCAGATAGCACAGG 120
DB 61 CCAGTTGGAGAACCATGCCCTGAAGTCTCTACATCCCGATGAGCAGATAGCACAGG 120

QY 121 ACCTGAGATGGCGCGCGAGGGCTTTGGCTCTCGGGTTCAGCCCGCCAGGGCTCAC 180
DB 121 ACCTGAGATGGCGCGCGAGGGCTTTGGCTCTCGGGTTCAGCCCGCCAGGGCTCAC 180

QY 181 TGTGGCAGCGGGGCGCCAGCCAAAGCAGCAGCAAGTGGATCCCTCTCGGCTCTCTGT 240
DB 181 TGTGGCAGCGGGGCGCCAGCCAAAGCAGCAGCAAGTGGATCCCTCTCGGCTCTCTGT 240

QY 241 GCCACCCAGTATGTGGTTCGCAATTTAGCAAGAGGGGGCCACCATCGGCAACATCAC 300
DB 241 GCCACCCAGTATGTGGTTCGCAATTTAGCAAGAGGGGGCCACCATCGGCAACATCAC 300

QY 301 ABAACAGACCCAGTCCAGATGACGTGATAGCAAGAGGAGAGCGAGTGCAGCTGAAAA 360
DB 301 ABAACAGACCCAGTCCAGATGACGTGATAGCAAGAGGAGAGCGAGTGCAGCTGAAAA 360

QY 361 AGCCATCAGTGTGCACTCCACCCCTGAGGCTCTCTCCGCTTGAAGATGATCTTGA 420
DB 361 AGCCATCAGTGTGCACTCCACCCCTGAGGCTCTCTCCGCTTGAAGATGATCTTGA 420

QY 421 GATTATGATAAGAGGCTTAAGACACCAAAAGCGCTGACGAGTTCCCTGAAATCTCT 480
DB 421 GATTATGATAAGAGGCTTAAGACACCAAAAGCGCTGACGAGTTCCCTGAAATCTCT 480

QY 481 GGCCCATATAAATTTGTAGGGCTCTCAITGGCAAGAGGAGGACGAACTTGAAGAAGT 540
DB 481 GGCCCATATAAATTTGTAGGGCTCTCAITGGCAAGAGGAGGACGAACTTGAAGAAGT 540

QY 541 AGAGCAAGATACCGAGCAAAAATACCATCTCTCTGTTGCAAGACCTTACCTTTACAA 600
DB 541 AGAGCAAGATACCGAGCAAAAATACCATCTCTCTGTTGCAAGACCTTACCTTTACAA 600

QY 601 CCCTGAGAGGACCATCACTGTGAAGGGGCGCATCGAGATTTGTCAGGGCGGAGCAGGA 660
DB 601 CCCTGAGAGGACCATCACTGTGAAGGGGCGCATCGAGATTTGTCAGGGCGGAGCAGGA 660

QY 661 AATAATGAAGAAAGTTGCGGAGGCTTATGAGATGATGTGGTGCATGAGCTCTCACCT 720
DB 661 AATAATGAAGAAAGTTGCGGAGGCTTATGAGATGATGTGGTGCATGAGCTCTCACCT 720

QY 721 GATCCCTGGCTGAACCTGGCTCTGTAGTCTTTTCCAGCTTTCATCAGGGCAGTCC 780
DB 721 GATCCCTGGCTGAACCTGGCTCTGTAGTCTTTTCCAGCTTTCATCAGGGCAGTCC 780

QY 781 GCCGCTCCAGCAGGCTTACTGGGCTGTCTCCCTATAGCTCTTTATGAGGCTCCCGA 840
DB 781 GCCGCTCCAGCAGGCTTACTGGGCTGTCTCCCTATAGCTCTTTATGAGGCTCCCGA 840

QY 841 GCAGGAGATGCTGAGGTTTATCCCGCCAGGAGTGGGGCCATCATCGGCAAGAA 900
DB 841 GCAGGAGATGCTGAGGTTTATCCCGCCAGGAGTGGGGCCATCATCGGCAAGAA 900

QY 901 GGGGCGAGCACATCAAAAGCTCTCCCGGTTTGGCAGCGCTCCATCAAGATTCACCCACC 960
DB 901 GGGGCGAGCACATCAAAAGCTCTCCCGGTTTGGCAGCGCTCCATCAAGATTCACCCACC 960

DB 901 GGGGCGAGCACATCAAAAGCTCTCCCGGTTTGGCAGCGCTCCATCAAGATTCACCCACC 960
QY 961 CGAAACACCTGACTCCAAAGTTCTATGTTATCATCTGGACCGCCAGAGGCCCAATT 1020
DB 961 CGAAACACCTGACTCCAAAGTTCTATGTTATCATCTGGACCGCCAGAGGCCCAATT 1020

QY 1021 CAAGGCTCAGGGAAGATCTATGCAAACTCAAGGAGGAGAACTTCTTTGGTCCCAAGGA 1080
DB 1021 CAAGGCTCAGGGAAGATCTATGCAAACTCAAGGAGGAGAACTTCTTTGGTCCCAAGGA 1080

QY 1081 GGAAGTGAAGCTGGAGACCCACATACGTTGTCAGAGCATCAGCAGCTGCCCGGGTCATTGG 1140
DB 1081 GGAAGTGAAGCTGGAGACCCACATACGTTGTCAGAGCATCAGCAGCTGCCCGGGTCATTGG 1140

QY 1141 CAAGGTTGGAAAAACGGTGAACCGAGTTGAGAAATTTGACGGCAGCTGAGGTGAGTACC 1200
DB 1141 CAAGGTTGGAAAAACGGTGAACCGAGTTGAGAAATTTGACGGCAGCTGAGGTGAGTACC 1200

QY 1201 AAGAGACAGACCCCTGATGAGAACGACAGGTCTATCGTGAATAATCATCGACATTTCTA 1260
DB 1201 AAGAGACAGACCCCTGATGAGAACGACAGGTCTATCGTGAATAATCATCGACATTTCTA 1260

QY 1261 TGCAGTCAAGTGGCTCAACGGAGATCCAGACATCTTGGCCAGGTTAAGCAGCAGCA 1320
DB 1261 TGCAGTCAAGTGGCTCAACGGAGATCCAGACATCTTGGCCAGGTTAAGCAGCAGCA 1320

QY 1321 TCAGAGGGACAGATTAACGAGGCCAGGACCGAGGAGGAGTACCAGCCCTCCCTGTCC 1380
DB 1321 TCAGAGGGACAGATTAACGAGGCCAGGACCGAGGAGGAGTACCAGCCCTCCCTGTCC 1380

QY 1381 CTGAGTCCAGGACAAACCGGCGAGAAATCGAGAGTGTCTCTCCCGGCGAGGCTGA 1440
DB 1381 CTGAGTCCAGGACAAACCGGCGAGAAATCGAGAGTGTCTCTCCCGGCGAGGCTGA 1440

QY 1441 GAATGAGTGGGATCCGGGACACNTGGCCGGGCTGTAGATCAGGTTTGGCCACTTGATT 1500
DB 1441 GAATGAGTGGGATCCGGGACACNTGGCCGGGCTGTAGATCAGGTTTGGCCACTTGATT 1500

QY 1501 GAGAAAGATGTTCCAGTGAGGAAACCTGTATCTNTCAGCCCCCAACACCCCAATTGGC 1560
DB 1501 GAGAAAGATGTTCCAGTGAGGAAACCTGTATCTNTCAGCCCCCAACACCCCAATTGGC 1560

QY 1561 CCAACACTGNTGCCCCCTCGGGTGTGAGAAATTTAGCGCAAGGCACTTTTAAACGTGG 1620
DB 1561 CCAACACTGNTGCCCCCTCGGGTGTGAGAAATTTAGCGCAAGGCACTTTTAAACGTGG 1620

QY 1621 ATTCTTTAAAGAGCTCTCCAGGCCCCCAAGAGGAGTGGATCACCTCAGTGGGAAGA 1680
DB 1621 ATTCTTTAAAGAGCTCTCCAGGCCCCCAAGAGGAGTGGATCACCTCAGTGGGAAGA 1680

QY 1681 AAAATAAAATTTCTTCAGGTTTAAAA 1708
DB 1681 AAAATAAAATTTCTTCAGGTTTAAAA 1708

RESULT 2
AR343073
LOCUS AR343073 1708 bp DNA linear PAT 17-AUG-2003
DEFINITION Sequence 5 from patent US 6576756.
ACCESSION AR343073
VERSION AR343073.1 GI:33738475
KEYWORDS Unknown.
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1708)
AUTHORS Chen,Y.-T., Gure,A., Tsang,S., Stockert,E., Jager,E., Alexander,K.
and Old,L.J.
TITLE Isolated nucleic acid molecule encoding cancer associated antigen,
the antigen itself, and uses thereof
JOURNAL Patent: US 6576756-A 5 10-JUN-2003;
FEATURES Location/Qualifiers
source 1..1708


```
/organism="unknown"  
/mol_type="genomic DNA"
```

```
Query Match      100.0%; Score 1708; DB 6; Length 1708;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1708; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Qy	1	AGGACCGTGC	CGCACCGCCCGCAGT	TTATCCCGGGAGCCATCATGAAGCTGAATGGCCA	60
Db	1	AGGACCGTGC	CGCACCGCCCGCAGT	TTATCCCGGGAGCCATCATGAAGCTGAATGGCCA	60
Qy	61	CCAGTTGGAGAA	CATGTCCTGGAAGTCTCCTACATCCCGATGAGCAGATAGCACAGG	120	
Db	61	CCAGTTGGAGAA	CATGTCCTGGAAGTCTCCTACATCCCGATGAGCAGATAGCACAGG	120	
Qy	121	ACCTGAGAAT	GGGCGCGAGGGGCTTTGGCTCTCGGGGT	CAGCCCGCCAGGGCTCAC	180
Db	121	ACCTGAGAAT	GGGCGCGAGGGGCTTTGGCTCTCGGGGT	CAGCCCGCCAGGGCTCAC	180
Qy	181	TGTGGCAGGGG	CGCCCGCAGCAAGCAGCAAGTGGA	CATCCCTTTGGCTCTGGT	240
Db	181	TGTGGCAGGGG	CGCCCGCAGCAAGCAGCAAGTGGA	CATCCCTTTGGCTCTGGT	240
Qy	241	GCCCAACCAT	GATGTGGGTGCCATTATTGGCAAGAGGGGCCACCATCCGCAACATCAC	300	
Db	241	GCCCAACCAT	GATGTGGGTGCCATTATTGGCAAGAGGGGCCACCATCCGCAACATCAC	300	
Qy	301	AAAACAGAC	CCAGTCCAAGATAGACGTGCATAGGAAGGAGAA	CGAGTGCAGCTGAAAA	360
Db	301	AAAACAGAC	CCAGTCCAAGATAGACGTGCATAGGAAGGAGAA	CGAGTGCAGCTGAAAA	360
Qy	361	AGCCATCAG	TGTGCACCTCCAGCCCTGAGGGTGTCTCTCCGTTGTAA	GATGATCTTGGA	420
Db	361	AGCCATCAG	TGTGCACCTCCAGCCCTGAGGGTGTCTCTCCGTTGTAA	GATGATCTTGGA	420
Qy	421	GATTATGCAT	AAAGAGGCTAAGGACACCAAAACGGCTGACGAGGT	TCCCTGGAAGATCCT	480
Db	421	GATTATGCAT	AAAGAGGCTAAGGACACCAAAACGGCTGACGAGGT	TCCCTGGAAGATCCT	480
Qy	481	GGGCCATAA	TAACTTTGTAGGGCGTCTCATTTGGCAAGGAAGGACGGA	ACCTGAAGAAGGT	540
Db	481	GGGCCATAA	TAACTTTGTAGGGCGTCTCATTTGGCAAGGAAGGACGGA	ACCTGAAGAAGGT	540
Qy	541	AGACRAGAT	ACCGAGCAAAATACCATCTCCTCGTTGCAAGACCT	TACCCCTTACAA	600
Db	541	AGACRAGAT	ACCGAGCAAAATACCATCTCCTCGTTGCAAGACCT	TACCCCTTACAA	600
Qy	601	CCCTGAGAG	ACCATCACTGTGAAGGGGCCATCGAGAA	TGTTGTCAGGGCCGAGCAGGA	660
Db	601	CCCTGAGAG	ACCATCACTGTGAAGGGGCCATCGAGAA	TGTTGTCAGGGCCGAGCAGGA	660
Qy	661	AATAATGA	AGAAAGTTCCGGAGGCCTATGAGAA	TGATGTGCTCCCATGAGCTCTCACCT	720
Db	661	AATAATGA	AGAAAGTTCCGGAGGCCTATGAGAA	TGATGTGCTCCCATGAGCTCTCACCT	720
Qy	721	GATCCCTG	SGCTGAACTGCTGTAGGTCTCTTTCCAGCTTCAT	CAGCGAGTCCC	780
Db	721	GATCCCTG	SGCTGAACTGCTGTAGGTCTCTTTCCAGCTTCAT	CAGCGAGTCCC	780
Qy	781	GC	CGCTCCCAGCAGCGTTACTGGGGTGTCTTATAGCT	TCTTTATGAGGGTCCCGA	840
Db	781	GC	CGCTCCCAGCAGCGTTACTGGGGTGTCTTATAGCT	TCTTTATGAGGGTCCCGA	840
Qy	841	GCAGGAGAT	GGTGAGGTGTTATCCCGGCCAGGCA	GTGGGCCCATCATCGCAAGAA	900
Db	841	GCAGGAGAT	GGTGAGGTGTTATCCCGGCCAGGCA	GTGGGCCCATCATCGCAAGAA	900
Qy	901	GGGSCAC	CATCAAAACAGCTCTCCCGTTGCGAGCGCTCCAT	CAAGATTGCACACC	960
Db	901	GGGSCAC	CATCAAAACAGCTCTCCCGTTGCGAGCGCTCCAT	CAAGATTGCACACC	960
Qy	961	CGAAAC	CCCTGACTCCAAAGTTCGTATGGTTATCATCACTGGAC	CGCGCAGAGGCCCAATT	1020

D5	961	CGAAACACCTGACTCCAAGATTCGTATGGTTATCATCTGGACCGCCAGAGCCCAATT	1020
QY	1021	CAAGGCTCAGGGAAGATCTATGCGAAATCAAGGAGGAGAACTCTTTTGGTCCCAAGGA	1080
D5	1021	CAAGGCTCAGGGAAGATCTATGCGAAATCAAGGAGGAGAACTCTTTTGGTCCCAAGGA	1080
QY	1081	GGAAAGTGAAGCTGGAGACCCACATACGTGTGCAGCATCAGCAGCTGGCCGGGTCAATTGG	1140
D5	1081	GGAAAGTGAAGCTGGAGACCCACATACGTGTGCAGCATCAGCAGCTGGCCGGGTCAATTGG	1140
QY	1141	CAAGGTGGAAAAACGGTGAACAGATTGCAGAAATTTGACGGCAGCTGAGGTGGTAGTACC	1200
D5	1141	CAAGGTGGAAAAACGGTGAACAGATTGCAGAAATTTGACGGCAGCTGAGGTGGTAGTACC	1200
QY	1201	AAGAGACCGACACCCCTGATCAGAGAACGACCAAGCTCATCTGCAAAATCATCGGACATTTCTA	1260
D5	1201	AAGAGACCGACACCCCTGATCAGAGAACGACCAAGCTCATCTGCAAAATCATCGGACATTTCTA	1260
QY	1261	TGCCAGTCAGATGGCTCAACCGGAAGATCCGAGACATCCTGGCCCAAGCTTTAAGCAGCAGCA	1320
D5	1261	TGCCAGTCAGATGGCTCAACCGGAAGATCCGAGACATCCTGGCCCAAGCTTTAAGCAGCAGCA	1320
QY	1321	TCAGAGGGGACAGATGAACCCAGGCCACAGGCACGGAGGAAGTGACACGACCCCTCCCTGTGTC	1380
D5	1321	TCAGAGGGGACAGATGAACCCAGGCCACAGGCACGGAGGAAGTGACACGACCCCTCCCTGTGTC	1380
QY	1381	CTTNGAGTCCAGGACAAACACGGGCGGAAATCGAGAGTGTGCTCTCCCGCGCAGGCGCTGA	1440
D5	1381	CTTNGAGTCCAGGACAAACACGGGCGGAAATCGAGAGTGTGCTCTCCCGCGCAGGCGCTGA	1440
QY	1441	GAATGAGTGGGAATCCGGGACACNTGGCCGGGCTGTAGTCAGGTTTCGCCACCTTGATT	1500
D5	1441	GAATGAGTGGGAATCCGGGACACNTGGCCGGGCTGTAGTCAGGTTTCGCCACCTTGATT	1500
QY	1501	GAGAAAGATGTTCCAGTGAGGAACCCCTGTATCTNTCAGCCCCCAACACCCCAATGGC	1560
D5	1501	GAGAAAGATGTTCCAGTGAGGAACCCCTGTATCTNTCAGCCCCCAACACCCCAATGGC	1560
QY	1561	CCAACTGNTGCCCTCGGGGTGTTCAGAAATNTAGCGCAAGGACATTTTAAACGTGG	1620
D5	1561	CCAACTGNTGCCCTCGGGGTGTTCAGAAATNTAGCGCAAGGACATTTTAAACGTGG	1620
QY	1621	ATTGTTTAAAGAGCTCTCCAGGCCCCACCAAGAGGGTGGATCACACCTCAGTGGGAAGA	1680
D5	1621	ATTGTTTAAAGAGCTCTCCAGGCCCCACCAAGAGGGTGGATCACACCTCAGTGGGAAGA	1680
QY	1681	AAAATAAAATTTCTTCAGGTTTTTAAA	1708
D5	1681	AAAATAAAATTTCTTCAGGTTTTTAAA	1708

RESULT	3
BD209924	
LOCUS	linear PAT 17-JUL-2003
DEFINITION	Isolated nucleic acid molecules encoding cancer-associated antigens, these antigens and method of using the same.
ACCESSION	BD209924
VERSION	BD209924.1 GI:33019694
KEYWORDS	JP 2002S12049-A/3.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 1708) Chen,Y.T., Gure,A., Tsang,S., Stockert,E., Jager,E., Knuth,A. and Old,J.J.
TITLE	Isolated nucleic acid molecules encoding cancer-associated antigens, these antigens and method of using the same
JOURNAL	Patent: JP 2002S12049-A 3 23-APR-2002;
COMMENT	LUDWIG INSTITUTE FOR CANCER RESEARCH OS Homo sapiens (human) PN JP 2002S12049-A/3

```

PD 23-APR-2002
PF 16-MAR-1999 JP 2000545030
PR 17-APR-1998 US 09/061709
PI YAO TSENG CHEN,ALI GURE,SOLAM TSANG,ELISABETH STOCKERT,ELKE
PI JAGER.
PI ALEXANDER KNUTH,LLOYD J OLD
PC C12N15/09,A61K35/12,A61K39/00,A61K39/39,A61P35/00,C07K16/32,
PC C12N1/15,
PC C12N1/19,C12N1/21,C12N5/10,C12P21/08,C12Q1/68,G01N33/53,G01N33/ PC
574,
PC C12N15/00,C12N5/00
CC n=not determined(1384,1464,1533,1571,1595)
FH Key Location/Qualifiers
FT source 1..1708
FT source /organism="Homo sapiens (human)".
FEATURES
source
1..1708
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 100.0%; Score 1708; DB 6; Length 1708;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1708; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AGGAGCGTCCCGACGGCCCGAGTTACCCCGGGAGGCATCATGAAGCTGAATGGCCA 60
DB 1 AGGAGCGTCCCGACGGCCCGAGTTACCCCGGGAGGCATCATGAAGCTGAATGGCCA 60
QY 61 CCAAGTTGGAGAACCATGCTGAAGCTCTCTACATCCCGGATGACGATAGCACAGGG 120
DB 61 CCAAGTTGGAGAACCATGCTGAAGCTCTCTACATCCCGGATGACGATAGCACAGGG 120
QY 121 ACTGAGAAATGGCGCGGAGGGGCTTTGGCTCTCGGGTTCAGCCCGCCAGGGCTCAC 180
DB 121 ACTGAGAAATGGCGCGGAGGGGCTTTGGCTCTCGGGTTCAGCCCGCCAGGGCTCAC 180
QY 181 TGTGGCAGCGGGGCGCCAGCAGAGCAGCAAGTGGAATCCCGCTTCGGCTCTCTGT 240
DB 181 TGTGGCAGCGGGGCGCCAGCAGAGCAGCAAGTGGAATCCCGCTTCGGCTCTCTGT 240
QY 241 GCCACCCAGTATGTGGTGCCATTATTTGGCAAGAGGGGGCCACCATCCGCAACATCAC 300
DB 241 GCCACCCAGTATGTGGTGCCATTATTTGGCAAGAGGGGGCCACCATCCGCAACATCAC 300
QY 301 AAAACAGACCCAGTCCAGATAGACCTGATAGGAGGAGAGCGAGGTGCAGCTCAAAA 360
DB 301 AAAACAGACCCAGTCCAGATAGACCTGATAGGAGGAGAGCGAGGTGCAGCTCAAAA 360
QY 361 AGCCATCAGTGTGCATCCACCCCTGAGGGCTGCTCTCCGCTTGTAAAGATGATCTTGA 420
DB 361 AGCCATCAGTGTGCATCCACCCCTGAGGGCTGCTCTCCGCTTGTAAAGATGATCTTGA 420
QY 421 GATTATGCATAAAGAGGCTTAAGACACAAAACGGCTGACGAGGTTCCTTGAAGATCT 480
DB 421 GATTATGCATAAAGAGGCTTAAGACACAAAACGGCTGACGAGGTTCCTTGAAGATCT 480
QY 481 GGCCCAATAAATTTGTAGGGGCTCTCATTTGCAAGGAGGACGGAACCTGAAAGAGGT 540
DB 481 GGCCCAATAAATTTGTAGGGGCTCTCATTTGCAAGGAGGACGGAACCTGAAAGAGGT 540
QY 541 AGAGCAAGATPACCGAGACAAAATACCATCTCTCTGTTGCAAGCTTACCTTTTACAA 600
DB 541 AGAGCAAGATPACCGAGACAAAATACCATCTCTCTGTTGCAAGCTTACCTTTTACAA 600
QY 601 CCCTGAGAGGACCATCATCTGTGAAGGGGCCATCGAGAAATGTTGACGAGGCGAGCAGGA 660
DB 601 CCCTGAGAGGACCATCATCTGTGAAGGGGCCATCGAGAAATGTTGACGAGGCGAGCAGGA 660
QY 661 AATAATGAAGAAAGTTGGGAGGCGCTATGAGAAATGATGTGGCTGCATGAGCTCTCACT 720

```

LOCUS AR171866 1946 bp DNA linear PAT 17-DEC-2001
DEFINITION Sequence 7 from patent US 6297364.
ACCESSION AR171866
VERSION AR171866.1 GI:17910816
KEYWORDS Unknown.
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1946)
AUTHORS Chen, Y.-T., Gure, A., Tsang, S., Stockert, E., Jager, E., Alexander, K. and Old, L. J.
TITLE Isolated nucleic acid molecule encoding cancer associated antigen, the antigen itself, and uses thereof
JOURNAL Patent: US 6297364-A 7 02-OCT-2001;
FEATURES Location/Qualifiers
source 1..1946
/organism="unknown"
/mol_type="unassigned DNA"
ORIGIN
Query Match 97.9%; Score 1672; DB 6; Length 1946;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1672; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 37 AGCCATCATGAAGCTGAATGGCCACCAGTGTGGAGAACCATGCCCTGAAGCTCTCTACAT 96
DB 275 AGCCATCATGAAGCTGAATGGCCACCAGTGTGGAGAACCATGCCCTGAAGCTCTCTACAT 334
QY 97 CCCCGATGACAGATAGCACAGGACCTGAGAAATGGCGCGCGAGGGGGCTTTGGCTCTCG 156
DB 335 CCCCGATGACAGATAGCACAGGACCTGAGAAATGGCGCGCGAGGGGGCTTTGGCTCTCG 394
QY 157 GGGTCAGCCCCCGCAGGGCTCACTGTGGCAGCGGGGGCCGACGACGAGCAAGT 216
DB 395 GGGTCAGCCCCCGCAGGGCTCACTGTGGCAGCGGGGGCCGACGACGAGCAAGT 454
QY 217 GGACATCCCCCTTCGGCTCTCGTGTGCCACCCAGTATGTGGTGCCATTATTGGCAAGGA 276
DB 455 GGACATCCCCCTTCGGCTCTCGTGTGCCACCCAGTATGTGGTGCCATTATTGGCAAGGA 514
QY 277 GGGGGCCACCATCGGCAACATCAAAACAGACCCAGTCCAAGATAGAGCTGCATAGGAA 336
DB 515 GGGGGCCACCATCGGCAACATCAAAACAGACCCAGTCCAAGATAGAGCTGCATAGGAA 574
QY 337 GGAGAACGCGAGTGCAGTGAAGAGCCATCAGTGTGCCTCCACCCCTGAGGGCTGCTC 396
DB 575 GGAGAACGCGAGTGCAGTGAAGAGCCATCAGTGTGCCTCCACCCCTGAGGGCTGCTC 634
QY 397 CTCGGCTTGTAAGATGATCTTGAGATTATGCAATAAGAGGGCTAAGGACACCAAAACGGC 456
DB 635 CTCGGCTTGTAAGATGATCTTGAGATTATGCAATAAGAGGGCTAAGGACACCAAAACGGC 694
QY 457 TGACGAGTTCCTCGAAGTCTGGGCCCATTAATTAATCTTTAGGGCGTCTCATTTGGGAA 516
DB 695 TGACGAGTTCCTCGAAGTCTGGGCCCATTAATTAATCTTTAGGGCGTCTCATTTGGGAA 754
QY 517 GGAAGACGGAACTCTGAAGAGGTAGAGCAAGATACCGAGACAAATAATCACCATCTCTC 576
DB 755 GGAAGACGGAACTCTGAAGAGGTAGAGCAAGATACCGAGACAAATAATCACCATCTCTC 814
QY 577 GTTGCAAGACCTTACCTTTAAACCTTGAGAGGACCATCACTGTGAAGGGGGCCATCGA 636
DB 815 GTTGCAAGACCTTACCTTTAAACCTTGAGAGGACCATCACTGTGAAGGGGGCCATCGA 874
QY 637 GAATTGTGTCAGGGCCGAGCAGGAATAATGAGAAAGTTCGGAGGGCTATGAGAAATGA 696
DB 875 GAATTGTGTCAGGGCCGAGCAGGAATAATGAGAAAGTTCGGAGGGCTATGAGAAATGA 934
QY 697 TGTGGCTGCATCAGCTCTCACTGATCCCTGGCCCTGAACCTGGCTGTGTAGGCTTTT 756
DB 935 TGTGGCTGCATCAGCTCTCACTGATCCCTGGCCCTGAACCTGGCTGTGTAGGCTTTT 994
QY 757 CCCAGCTTCATCAGCGCAGTCCCGCGGCTCCAGCAGGGTTACTGGGGCTGCTCCCTA 816
DB 995 CCCAGCTTCATCAGCGCAGTCCCGCGGCTCCAGCAGGGTTACTGGGGCTGCTCCCTA 1054
QY 817 TAGTCTCTTTATGAGGCTCCCGAGCAGAGATGGTGCAGGTGTTTATTCGCCGCCAGGC 876
DB 1055 TAGTCTCTTTATGAGGCTCCCGAGCAGAGATGGTGCAGGTGTTTATTCGCCGCCAGGC 1114
QY 877 AGTGGGCGCCCATCATCGGCAAGAGGGGCGACACATCAAAACAGCTCTCCCGGTTGGCAG 936
DB 1115 AGTGGGCGCCCATCATCGGCAAGAGGGGCGACACATCAAAACAGCTCTCCCGGTTGGCAG 1174
QY 937 CGCTCCCATCAAGATTGCAACCCGAAACCTGACTTCCAAAGTTCGTATGGTTATCAT 996
DB 1175 CGCTCCCATCAAGATTGCAACCCGAAACCTGACTTCCAAAGTTCGTATGGTTATCAT 1234
QY 997 CACTGGACCGCAGAGGCCCAATCAAGGCTCAGGAGAGATCTATGCAAACTCAAGGA 1056
DB 1235 CACTGGACCGCAGAGGCCCAATCAAGGCTCAGGAGAGATCTATGCAAACTCAAGGA 1294
QY 1057 GGAGAACTTTCTTTGGTCCCAAGAGGAGTGAAGCTGGAGACCCACATACGTGTGCCAGC 1116
DB 1295 GGAGAACTTTCTTTGGTCCCAAGAGGAGTGAAGCTGGAGACCCACATACGTGTGCCAGC 1354
QY 1117 ATCAGCAGCTGGCGGGTCAATGGCAAGGTGGAAACCGTGAACGATTCGAGAAATTT 1176
DB 1355 ATCAGCAGCTGGCGGGTCAATGGCAAGGTGGAAACCGTGAACGATTCGAGAAATTT 1414
QY 1177 GACGCGAGCTGAGGTGTAGTACCAAGAGACCAAGCCCTCATGAGAACGACACAGGTGAT 1236
DB 1415 GACGCGAGCTGAGGTGTAGTACCAAGAGACCAAGCCCTCATGAGAACGACACAGGTGAT 1474
QY 1237 CGTGAANAATCATCGGACATTTCTATGTCAGTCAAGTGGCTCAACGGAGATTCGAGAGAT 1296
DB 1475 CGTGAANAATCATCGGACATTTCTATGTCAGTCAAGTGGCTCAACGGAGATTCGAGAGAT 1534
QY 1297 CCGTGGCCAGGTTAAGCAGCAGCATCAGAGGGACAGATTAACGAGCCACAGGCGCCAGG 1356
DB 1535 CCGTGGCCAGGTTAAGCAGCAGCATCAGAGGGACAGATTAACGAGCCACAGGCGCCAGG 1594
QY 1357 GAAGTGAACACGCCCTCCCTGTCCCTTNGAGTCCAGGACAAACACGGGACAGAAATCGAGA 1416
DB 1595 GAAGTGAACACGCCCTCCCTGTCCCTTNGAGTCCAGGACAAACACGGGACAGAAATCGAGA 1654
QY 1417 GTGTGCTCTCCCGCAGGCTGAGATGAGTGGGAATCCGGGACACNTGGGCGGGCTG 1476
DB 1655 GTGTGCTCTCCCGCAGGCTGAGATGAGTGGGAATCCGGGACACNTGGGCGGGCTG 1714
QY 1477 TAGATCAGGTTTGCCCACTTGATTGAGAAAGATGTTCCAGTGAGGAAACCTTGATCTNTCA 1536
DB 1715 TAGATCAGGTTTGCCCACTTGATTGAGAAAGATGTTCCAGTGAGGAAACCTTGATCTNTCA 1774
QY 1537 GCGCCAAACACCCACCCAAATTTGGCCCAACATGNTGCCCTCGGGGTGTGAGAAATTT 1596
DB 1775 GCGCCAAACACCCACCCAAATTTGGCCCAACATGNTGCCCTCGGGGTGTGAGAAATTT 1834
QY 1597 AGCGCAAGGCACTTTTAAACGTGATGTTTAAAGAGTCTCCAGGCCCCACCAAGAGG 1656
DB 1835 AGCGCAAGGCACTTTTAAACGTGATGTTTAAAGAGTCTCTCCAGGCCCCACCAAGAGG 1894
QY 1657 GTGATCAGCCTCAGTGGGGAAGAAATAAATTTCTTCAGGTTTTTAAAA 1708
DB 1895 GTGATCAGCCTCAGTGGGGAAGAAATAAATTTCTTCAGGTTTTTAAAA 1946
RESULT 5
AR343075
LOCUS AR343075 1946 bp DNA linear PAT 17-AUG-2003
DEFINITION Sequence 7 from patent US 6576756.
ACCESSION AR343075
VERSION AR343075.1 GI:33738477
KEYWORDS Unknown.
SOURCE Unknown.
ORGANISM Unknown.

Unclassified.
 1 (bases 1 to 1946)
 Chen, Y.-T., Gure, A., Tsang, S., Stockert, E., Jager, E., Alexander, K.
 and Old, L.J.
 TITLE Isolated nucleic acid molecule encoding cancer associated antigen,
 the antigen itself, and uses thereof
 JOURNAL Patent: US 6576756-A 7 10-JUN-2003;
 FEATURES Location/Qualifiers
 source 1..1946
 /organism="unknown"
 /mol_type="genomic DNA"

ORIGIN

Query Match 97.9%; Score 1672; DB 6; Length 1946;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1672; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 37 AGCCATCATGAAGCTGAATGGCCACCAGTGGAGAACCATGCCCTGAAGGTCTCTACAT 96
 DB 275 AGCCATCATGAAGCTGAATGGCCACCAGTGGAGAACCATGCCCTGAAGGTCTCTACAT 334

QY 97 CCCGATGACGATAGCAGGAGCCTGAGAAATGGGCCCGAGGGGCTTTGGCTCTCG 156
 DB 335 CCCGATGACGATAGCAGGAGCCTGAGAAATGGGCCCGAGGGGCTTTGGCTCTCG 394

QY 157 GGGTCAGCCCCCGCAGGGCTCACTGTGGCAGGGGGGCCAGCAAGCAGCAGCAAGT 216
 DB 395 GGGTCAGCCCCCGCAGGGCTCACTGTGGCAGGGGGGCCAGCAAGCAGCAGCAAGT 454

QY 217 GGCATCTCCCTTCGGCTCTGGTCCACCCAGTATGTGGTGCCATATTATGGCAAGGA 276
 DB 455 GGCATCTCCCTTCGGCTCTGGTCCACCCAGTATGTGGTGCCATATTATGGCAAGGA 514

QY 277 GGGGGCCACCATCGCAATCAAAAAGAGCCAGTCCAGATAGAGCTGCATAGGAA 336
 DB 515 GGGGGCCACCATCGCAATCAAAAAGAGCCAGTCCAGATAGAGCTGCATAGGAA 574

QY 337 GGAGAACGCGAGTGCAGCTGAAAGGCCATCATAGTGTGCACTCCACCCCTGAGGGCTGCTC 396
 DB 575 GGAGAACGCGAGTGCAGCTGAAAGGCCATCATAGTGTGCACTCCACCCCTGAGGGCTGCTC 634

QY 397 CTCGGCTGTAGATGATCTTGAGATATGATCAAAAGAGGCTAGGACACCAAGCCGC 456
 DB 635 CTCGGCTGTAGATGATCTTGAGATATGATCAAAAGAGGCTAGGACACCAAGCCGC 694

QY 457 TGACGAGGTTCCCTCAAGATCTGGCCCATATATCACTTTGTAGGCGCTCTCATTTGGCAA 516
 DB 695 TGACGAGGTTCCCTCAAGATCTGGCCCATATATCACTTTGTAGGCGCTCTCATTTGGCAA 754

QY 517 GGAAGACGGAACCTGAGNAGTAGACAAGTACCGAGACAAAATCACCATCTCTCTC 576
 DB 755 GGAAGACGGAACCTGAGNAGTAGACAAGTACCGAGACAAAATCACCATCTCTCTC 814

QY 577 GTTTCGAAGACCTTACCCCTTTACAACTCTGAGAGGACCATCACTGTGAAGGGGGCCATCGA 636
 DB 815 GTTTCGAAGACCTTACCCCTTTACAACTCTGAGAGGACCATCACTGTGAAGGGGGCCATCGA 874

QY 637 GAATGTGTGAGGGCCGAGCAGGAATAATGAAGAAAGTTTCGGAGGCGCTATGAGAAAGA 696
 DB 875 GAATGTGTGAGGGCCGAGCAGGAATAATGAAGAAAGTTTCGGAGGCGCTATGAGAAAGA 934

QY 697 TGTGGCTGCCATGAGTCTCACTGATCCCTGGCTGAACTGGCTGCTGTAGGCTCTTTT 756
 DB 935 TGTGGCTGCCATGAGTCTCACTGATCCCTGGCTGAACTGGCTGCTGTAGGCTCTTTT 994

QY 757 CCAGGTTTCATCCAGGCGAGTCCCGCGCTCCAGCAGCGTTACTTGGGGCTGCTCCCTA 816
 DB 995 CCAGGTTTCATCCAGGCGAGTCCCGCGCTCCAGCAGCGTTACTTGGGGCTGCTCCCTA 1054

QY 817 TAGCTCTTTATCAGGCTCCCGAGCAGGAGATGGTGAGGCTTTATCCCGCCCGCAGGC 876
 DB 1055 TAGCTCTTTATCAGGCTCCCGAGCAGGAGATGGTGAGGCTTTATCCCGCCCGCAGGC 1114

QY 877 AGTGGCGCCCATCATCGGCAAGAGGGCGAGCACATCAACAGCTCTCCGGTTGCCAG 936
 DB 1115 AGTGGCGCCCATCATCGGCAAGAGGGCGAGCACATCAACAGCTCTCCGGTTGCCAG 1174

QY 937 CGCTCCATCAAGATTGACACCCCGAACAACCTGACTCCAAAGTTCTGATGTTATCAT 996
 DB 1175 CGCTCCATCAAGATTGACACCCCGAACAACCTGACTCCAAAGTTCTGATGTTATCAT 1234

QY 997 CACTGGACCCGAGAGGCCCAATTCAGGCTCAGGAGAGNATCTATGGCAACTCAAGA 1056
 DB 1235 CACTGGACCCGAGAGGCCCAATTCAGGCTCAGGAGAGNATCTATGGCAACTCAAGA 1294

QY 1057 GGAGAACTCTTTGGTCCCAAGGAGGAAGTGAAGCTGGAGACCCACATACGTGTGCCAG 1116
 DB 1295 GGAGAACTCTTTGGTCCCAAGGAGGAAGTGAAGCTGGAGACCCACATACGTGTGCCAG 1354

QY 1117 ATCAGCAGCTGGCGGGTCAATTGGCAAGGTGAAAAAACGGTGAAACGAGTTGCAGAAATTT 1176
 DB 1355 ATCAGCAGCTGGCGGGTCAATTGGCAAGGTGAAAAAACGGTGAAACGAGTTGCAGAAATTT 1414

QY 1177 GACGGCAGCTGAGTGTGTAGTACCAAGAGACCAAGCCCTGATGAGAACGACAGTGCAT 1236
 DB 1415 GACGGCAGCTGAGTGTGTAGTACCAAGAGACCAAGCCCTGATGAGAACGACAGTGCAT 1474

QY 1237 CGTGAATAATCATCGGACATTTCTATGCCAGTCAAGTGGTCAACGGAAGATCCGAGACAT 1296
 DB 1475 CGTGAATAATCATCGGACATTTCTATGCCAGTCAAGTGGTCAACGGAAGATCCGAGACAT 1534

QY 1297 CCTGGCCCGGTTAAGCAGCAGCATCAGAGGAGCAGAGTAAACAGGCCCGCAGCAGGAG 1356
 DB 1535 CCTGGCCCGGTTAAGCAGCAGCATCAGAGGAGCAGAGTAAACAGGCCCGCAGCAGGAG 1594

QY 1357 GAAAGTACCAGCCCTCCCTGTCCTTNGAGTCCAGACCAACCAACGGCGAGAAATCGAGA 1416
 DB 1595 GAAAGTACCAGCCCTCCCTGTCCTTNGAGTCCAGACCAACCAACGGCGAGAAATCGAGA 1654

QY 1417 GTGTGCTCTCCCGGAGGCTGAGATGATGAGTGGGATCCGGACACATNTGGGCGGGCTG 1476
 DB 1655 GTGTGCTCTCCCGGAGGCTGAGATGATGAGTGGGATCCGGACACATNTGGGCGGGCTG 1714

QY 1477 TAGATCAGGTTTGCCCACTTGTATTGAAAGATGTTCCAGTGAAGAACCTGATCTNTCA 1536
 DB 1715 TAGATCAGGTTTGCCCACTTGTATTGAAAGATGTTCCAGTGAAGAACCTGATCTNTCA 1774

QY 1537 GCCCAACACCCACCCCAATGGCCCAACACTGTNTGCCCTCGGGGTGTCAGAAATTTNT 1596
 DB 1775 GCCCAACACCCACCCCAATGGCCCAACACTGTNTGCCCTCGGGGTGTCAGAAATTTNT 1834

QY 1597 AGCGAAGGCACTTTTAAACGTGATTTTAAAGAGAGCTCTCCAGGCCCCACCAAGAGG 1656
 DB 1835 AGCGAAGGCACTTTTAAACGTGATTTTAAAGAGAGCTCTCCAGGCCCCACCAAGAGG 1894

QY 1657 GTGATCACAACCTCAGTGGGAAGAAAATAAATTTCTTCAGGTTTAAAA 1708
 DB 1895 GTGATCACAACCTCAGTGGGAAGAAAATAAATTTCTTCAGGTTTAAAA 1946

RESULT 6
 BD209926
 LOCUS Isolated nucleic acid molecules encoding cancer-associated
 DEFINITION antigens, these antigens and method of using the same.
 ACCESSION BD209926
 VERSION BD209926.1 GI:33019696
 KEYWORDS JP 2002512049-A/5.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 1946)
 AUTHORS Chen, Y.T., Gure, A., Tsang, S., Stockert, E., Jager, E., Knuth, A. and
 Old, L.J.
 TITLE Isolated nucleic acid molecules encoding cancer-associated

antigens, these antigens and method of using the same			
JOURNAL	Patent: JP 2002512049-A 5 23-APR-2002;		
COMMENT	LUDWIG INSTITUTE FOR CANCER RESEARCH		
OS	Homo sapiens (human)		
PN	JP 2002512049-A/5		
PD	23-APR-2002		
PF	16-MAR-1999 JP 2000545030		
PR	17-APR-1998 US 09/061709		
PI	YAO TSENG CHEN,ALI GURE,SOLAM TSANG,ELISABETH STOCKERT,ELKE JAGER,		
PI	ALEXANDER KNUTH,LLOYD J OLD		
PC	C12N15/09,A61K35/12,A61K39/00,A61K39/39,A61P35/00,C07K16/32,		
PC	C12N1/15,		
PC	C12N1/19,C12N1/21,C12N5/10,C12P21/08,C12Q1/68,G01N33/53,G01N33/PC		
PC	574,		
CC	C12N15/00,C12N5/00		
CC	n=not determined(1622,1702,1771,1809,1833)		
PH	Key Location/Qualifiers		
FT	source 1..1946		
FT	Location/Qualifiers /organism='Homo sapiens (human)'		
FEATURES	1..1946		
source	/organism="Homo sapiens"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:9606"		
ORIGIN			
Query Match 97.9%; Score 1672; DB 6; Length 1946;			
Best Local Similarity 100.0%; Pred. No. 0;			
Matches 1672; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
QY	37 AGCCATCATGAAGTGAATGGCCACAGTTGGAGAACCATGCCCTGAAGGTCTCTACAT	96	
DB	275 AGCCATCATGAAGTGAATGGCCACAGTTGGAGAACCATGCCCTGAAGGTCTCTACAT	334	
QY	97 CCCGATGAGCAGATAGCAGGAGACCTGAGAACTGGCGCGCGAGGGGGCTTTGGCTCTCG	156	
DB	335 CCCGATGAGCAGATAGCAGGAGACCTGAGAACTGGCGCGCGAGGGGGCTTTGGCTCTCG	394	
QY	157 GGTCTAGCCCGCCAGGGCTCAGTGTGGCAGCGGGGGCCCGACCGACGAGCAAGT	216	
DB	395 GGTCTAGCCCGCCAGGGCTCAGTGTGGCAGCGGGGGCCCGACCGACGAGCAAGT	454	
QY	217 GGACATCCCGCTTCGGCTCTCGTGGCCACCCAGTATGTGGGTGCCATTATTGGCAAGGA	276	
DB	455 GGACATCCCGCTTCGGCTCTCGTGGCCACCCAGTATGTGGGTGCCATTATTGGCAAGGA	514	
QY	277 GGGGGCCACCATCCGCAACATCAAAACAGACCCAGTCCAGATAGAGCTGCATAGGAA	336	
DB	515 GGGGGCCACCATCCGCAACATCAAAACAGACCCAGTCCAGATAGAGCTGCATAGGAA	574	
QY	337 GGAGAGCGAGGTGCAGTGAAGAGCCATCAGTGTGCATCCACCCCTGAGGGCTGCTC	396	
DB	575 GGAGAGCGAGGTGCAGTGAAGAGCCATCAGTGTGCATCCACCCCTGAGGGCTGCTC	634	
QY	397 CTCGGCTTGAAGATGATCTTGGAGATTATGCATAAAGAGGCTTAAGGACACCAAAACGGC	456	
DB	635 CTCGGCTTGAAGATGATCTTGGAGATTATGCATAAAGAGGCTTAAGGACACCAAAACGGC	694	
QY	457 TGACGAGGTTCCTCTGAGATCTTGGCCCATATAACTTTGTAGGGCGTCTCATTTGGGAA	516	
DB	695 TGACGAGGTTCCTCTGAGATCTTGGCCCATATAACTTTGTAGGGCGTCTCATTTGGGAA	754	
QY	517 GGAAGACGGAACTGGAAGAGGTAGAGCAAGATACCGAGACAAAAATCACCATCTCTCTC	576	
DB	755 GGAAGACGGAACTGGAAGAGGTAGAGCAAGATACCGAGACAAAAATCACCATCTCTCTC	814	
QY	577 GTTGCAAGACCTTACCCCTTTACACCTTGAGAGGACCATCATCTGTGAAGGGGGCCATCGA	636	
DB	815 GTTGCAAGACCTTACCCCTTTACACCTTGAGAGGACCATCATCTGTGAAGGGGGCCATCGA	874	
QY	637 GAATTGTTCAGGGCCGAGCAGGAATAATGAGAAAGTTCCGGAGGCCCTATGAGATGA	696	

RESULT 7

DB	875	GAATTGTTGCAGGGCCGAGCAGGAAATAATGAAGAAAGTTCCGGAGGCTATGAGAATGA	934
QY	697	TGTGCTGCCATGAGCTCTCACCTGATCCCTGGCTGAACCTGGCTGCTGTAGGTCTTTT	756
DB	935	TGTGCTGCCATGAGCTCTCACCTGATCCCTGGCTGAACCTGGCTGCTGTAGGTCTTTT	994
QY	757	CCAGCTTCATCCAGCGCAGTCCCGCCGCTCCAGCAGCGCTTACTGGGGTGTCCCTA	816
DB	995	CCAGCTTCATCCAGCGCAGTCCCGCCGCTCCAGCAGCGCTTACTGGGGTGTCCCTA	1054
QY	817	TAGTCTCTTTTATGAGGCTCCCGAGCAGGAGATGGTGCAGGTGTTATCCCGCCGAGGC	876
DB	1055	TAGTCTCTTTTATGAGGCTCCCGAGCAGGAGATGGTGCAGGTGTTATCCCGCCGAGGC	1114
QY	877	AGTGGCGCCCATCATCGCAAGAGGGGCGACATCAAAACAGCTCTCTCCCGTTTGCAG	936
DB	1115	AGTGGCGCCCATCATCGCAAGAGGGGCGACATCAAAACAGCTCTCTCCCGTTTGCAG	1174
QY	937	CGCTTCCATCAAGATTGCACCCCGAAGACCTGACTTCCAAAGTTGCTATGTTATCAT	996
DB	1175	CGCTTCCATCAAGATTGCACCCCGAAGACCTGACTTCCAAAGTTGCTATGTTATCAT	1234
QY	997	CACGTGACCGCGCAGAGGCCCAATTCAAGGCTCAGGGAAGAACTCTATGGCAAACTCAAGGA	1056
DB	1235	CACGTGACCGCGCAGAGGCCCAATTCAAGGCTCAGGGAAGAACTCTATGGCAAACTCAAGGA	1294
QY	1057	GGAGAACTTCTTGGTCCCAAGAGAGAGTGAAGCTGGAGACCCACATACCTGTGCGCAGC	1116
DB	1295	GGAGAACTTCTTGGTCCCAAGAGAGAGTGAAGCTGGAGACCCACATACCTGTGCGCAGC	1354
QY	1117	ATCAGCAGCTGGCGGGTCAATTGGCAAGGTGGAAAAACGGTGAACGAGTTGCGAATTT	1176
DB	1355	ATCAGCAGCTGGCGGGTCAATTGGCAAGGTGGAAAAACGGTGAACGAGTTGCGAATTT	1414
QY	1177	GACGCGAGCTGAGGTGGTAGTACCAAGAGACCGACCCCTGATGAGAACCGACAGGTCT	1236
DB	1415	GACGCGAGCTGAGGTGGTAGTACCAAGAGACCGACCCCTGATGAGAACCGACAGGTCT	1474
QY	1237	CGTGAANAATCATCGACATTTCTATGCCAGTCAGATGGCTCAAAGGAGATCCGAGACAT	1296
DB	1475	CGTGAANAATCATCGACATTTCTATGCCAGTCAGATGGCTCAAAGGAGATCCGAGACAT	1534
QY	1297	CCTGGCCAGGTTAAGCAGCAGCATCAAGAGGACAGAGTAACCGAGGCCAGGACCGAG	1356
DB	1535	CCTGGCCAGGTTAAGCAGCAGCATCAAGAGGACAGAGTAACCGAGGCCAGGACCGAG	1594
QY	1357	GAAGTGACAGCCCTCTCCTCTTNGAGTCCAGAGCAACAAACGGGCGAGAAATCGAGA	1416
DB	1595	GAAGTGACAGCCCTCTCCTCTTNGAGTCCAGAGCAACAAACGGGCGAGAAATCGAGA	1654
QY	1417	GTGTGCTCTCCCGCAGGCTGTAGAAATGAGTGGGAATCCGGGACACNTGGCCGGCTG	1476
DB	1655	GTGTGCTCTCCCGCAGGCTGTAGAAATGAGTGGGAATCCGGGACACNTGGCCGGCTG	1714
QY	1477	TAGATCAGGTTTGGCCACTTGAATTGAGAAAGATGTTCCAGTGAGAAACCCCTGATCTNTCA	1536
DB	1715	TAGATCAGGTTTGGCCACTTGAATTGAGAAAGATGTTCCAGTGAGAAACCCCTGATCTNTCA	1774
QY	1537	GCCCCAACACCCCAANTTGGCCCACTGTTGCCCCCTCGGGTGTGAGAAATNT	1596
DB	1775	GCCCCAACACCCCAANTTGGCCCACTGTTGCCCCCTCGGGTGTGAGAAATNT	1834
QY	1597	AGCGCAAGGCACTTTTAAACCGTGGATTGTTTAAAGAAAGCTCTCCAGGCCCCCAGGAGG	1656
DB	1835	AGCGCAAGGCACTTTTAAACCGTGGATTGTTTAAAGAAAGCTCTCCAGGCCCCCAGGAGG	1894
QY	1657	GTGATCACACCTCAGTGGGGAAGAAAATAAAATTTCTTCAGGTTTAAAA	1708
DB	1895	GTGATCACACCTCAGTGGGGAAGAAAATAAAATTTCTTCAGGTTTAAAA	1946

```

AF117106      2130 bp  mRNA  linear  PRI 26-JAN-1999
LOCUS      Homo sapiens IGF-II mRNA-binding protein 1 (IMP-1) mRNA, complete
DEFINITION
cds.
ACCESSION  AF117106
VERSION    AF117106.1
KEYWORDS   GI:4191607
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 2130)
AUTHORS    Nielsen,J., Christiansen,J., Lykke-Andersen,J., Johnsen,A.H.,
Wewer,U.M. and Nielsen,F.C.
TITLE      A family of insulin-like growth factor II mRNA-binding proteins
represses translation in late development
JOURNAL    Mol. Cell. Biol. 19 (2), 1262-1270 (1999)
MEDLINE    99108099
PUBMED     9891060
REFERENCE   2 (bases 1 to 2130)
AUTHORS    Nielsen,J., Christiansen,J., Lykke-Andersen,J., Johnsen,A.H.,
Wewer,U.M. and Nielsen,F.C.
TITLE      Direct Submission
JOURNAL    Submitted (30-DEC-1998) Institute of Molecular Biology, University
of Copenhagen, Soelvgade 83H, Copenhagen DK-1307, Denmark
FEATURES   Location/Qualifiers
            source          1..2130
                        /organism="Homo sapiens"
                        /mol_type="mRNA"
                        /db_xref="taxon:9606"
            gene            1..2130
                        /gene="IMP-1"
            CDS              10..1743
                        /note="binds multiple sites in IGF-II mRNA 5'UTR"
                        /codon_start=1
                        /product="IGF-II mRNA-binding protein 1"
                        /protein_id="AAD09826.1"
                        /db_xref="GI:4191608"
                        /translation="NNKLVLGNLSVTPADLEKVPFAEHKISYGGFLVKSQYAFVDC
                        PHEWMAKAIETFSQKVELQGRLEIHSVPKQKSRKIQINIPQLRWELVDLLA
                        QGVTVNCSQVNSTEVAVVNTVYNSNREQTQALIMKLNGHLENHALKVSYPDEQIA
                        QPENGRRGFSQPGOPGSPVAAGAPAKQQQVDIPRLALVPTQYVGAIIKKEGATI
                        RNITKQTSKI DVHREKENAAGAAKATSVHSTPEGSSACKMILEIMHKEADTKTAD
                        VPLKILAHNNFVRLIGKGRNLKXVBODTEKTIISLQDLTLVNPRTITVKAIE
                        NCCRASQETMKVREAYENDVAAMSLQSHLIPGLNLAAGLFPASSSVAPPSPSVTG
                        AAPYSFPGAPQEMVQVPIPAQAVGAILGKGOHLKOLSRPASASIKIAPPEPDSK
                        VRNVITGPEAFKQAGRIYKGLKEENFFGPKREVKLETHIRVPASAGRVIGKGGK
                        TVNELQNLTAAEVVVPRDQTPDENQVIVKIIIGHFYASQMAQRKIRIDILAQVKQOHHQ
                        GGSNQARRK"
            ORIGIN
            Query Match      58.4%; Score 997; DB 9; Length 2130;
            Best Local Similarity 99.3%; Pred. No. 0;
            Matches 1667; Conservative 0; Mismatches 5; Indels 6; Gaps 1;

            QY 37 AGCCATCATGAAGCTGAATGGCCACCAGTTGGAGAACCATGCCTCGAAGGTCCTCTACAT 96
            DB 411 AGCCATCATGAAGCTGAATGGCCACCAGTTGGAGAACCATGCCTCGAAGGTCCTCTACAT 470
            QY 97 CCCCAGATGACATAGCAGCGACCTCAGATGGCGCCGAGGGGCTTTGGCTCTCG 156
            DB 471 CCCCAGATGACATAGCAGCGACCTCAGATGGCGCCGAGGGGCTTTGGCTCTCG 530
            QY 157 GGTCAGCCCCCGCAGGGCTCACTGTGGCAGCGGGGGCCCCAGCAAGCAGCAAGT 216
            DB 531 GGCTCAGCCCCCGCAGGGCTCACTGTGGCAGCGGGGGCCCCAGCAAGCAGCAAGT 590
            QY 217 GGCATCCCCCTTCGGCTCTGGTGGCCACCAGATGTGGTGCCATTTTGGCAAGGA 276
            DB 591 GGCATCCCCCTTCGGCTCTGGTGGCCACCAGATGTGGTGCCATTTTGGCAAGGA 650
            QY 277 GGGGGCCACCATCCGCAACATCACAAACAGACCCAGTCCTCAAGATAGACGTGCTAGGAA 336

```

```

651 GGGGGCCACCATCGCAACATCAAAACAGACCCAGTCCAAAGATAGACGTGATAGGAA 710
QY 337 GGAGAACGCGAGGTGACGTGAAAAGCCATCAGTGTGCACTCCACCCCTCAGGGCTGCTC 396
DB 711 GGAGAACGCGAGGTGACGTGAAAAGCCATCAGTGTGCACTCCACCCCTCAGGGCTGCTC 770
QY 397 CTCGCTTGTAAGATGATCTTGGAGATTATGCATAAAGAGGCTAAGGACACCAAAACGGC 456
DB 771 CTCGCTTGTAAGATGATCTTGGAGATTATGCATAAAGAGGCTAAGGACACCAAAACGGC 830
QY 457 TGACGAGGTCCCTGGAAGATCCTGGCCCATATAATCTTTGTAGGGGCTCTCAATTGGCAA 516
DB 831 TGACGAGGTCCCTGGAAGATCCTGGCCCATATAATCTTTGTAGGGGCTCTCAATTGGCAA 890
QY 517 GGAAGGACGGAACCTGGAAGAGGTAGACGAAAGATACCGAGACAAAATACCACTCTCCTC 576
DB 891 GGAAGGACGGAACCTGGAAGAGGTAGACGAAAGATACCGAGACAAAATACCACTCTCCTC 950
QY 577 GTTCCAGACCTTACCTTTTACAACCTGAGAGGACCATCATCTGTGAAGGGGCGCATCGA 636
DB 951 GTTCCAGACCTTACCTTTTACAACCTGAGAGGACCATCATCTGTGAAGGGGCGCATCGA 1010
QY 637 GAATTGTTGAGGGCCGAGCAGGAATAATGAAGAAAGTTTCGGGAGGCCCTATGAGAATGA 696
DB 1011 GAATTGTTGAGGGCCGAGCAGGAATAATGAAGAAAGTTTCGGGAGGCCCTATGAGAATGA 1070
QY 697 TGTGGCTGCCATGAGC-----TCTCACCTGATCCCTGGCCCTGAACCTGGCTCTCTAGG 750
DB 1071 TGTGGCTGCCATGAGCCTGCGAGTCTCACTGTATCCCTGGCTGAACCTGGCTCTCTAGG 1130
QY 751 TCTTTTCCAGCTTCATCCAGCGCAGTCCCGCGCGCTCCAGCAGCGGTACTCGGGCTGC 810
DB 1131 TCTTTTCCAGCTTCATCCAGCGCAGTCCCGCGCGCTCCAGCAGCGGTACTCGGGCTGC 1190
QY 811 TCCCTATAGTCTCTTTATGACGCTCCCGAGCAGGAGATGGTCAGGTGTTTATCCCCCG 870
DB 1191 TCCCTATAGTCTCTTTATGACGCTCCCGAGCAGGAGATGGTCAGGTGTTTATCCCCCG 1250
QY 871 CCAGCAGTGGCGCCCATCATCCGAGAGAGGGCGAGCATCAACACAGCTCTCCCGGT 930
DB 1251 CCAGCAGTGGCGCCCATCATCCGAGAGAGGGCGAGCATCAACACAGCTCTCCCGGT 1310
QY 931 TGCAGCGCCTCCATCAAGATTGACACCCGCGAAACACCTGACTCCAAAGTTCGTATGT 990
DB 1311 TGCCAGCGCCTCCATCAAGATTGACACCCGCGAAACACCTGACTCCAAAGTTCGTATGT 1370
QY 991 TATCATCACTGGACCGCAGAGGCCCAATTCAGAGGCTCAGGGAAGATCTATGGCAACT 1050
DB 1371 TATCATCACTGGACCGCAGAGGCCCAATTCAGAGGCTCAGGGAAGATCTATGGCAACT 1430
QY 1051 CAAGGAGGAGAACTTTCTTGGTCCCAAGGAGGAAGTGAAGCTGGAGACCCACATACGTGT 1110
DB 1431 CAAGGAGGAGAACTTTCTTGGTCCCAAGGAGGAAGTGAAGCTGGAGACCCACATACGTGT 1490
QY 1111 GCCAGCATCAGCAGCTGGCGGGTCAATTGSCAAAGGTGGAATAACCGTGAAACGAGTTGCA 1170
DB 1491 GCCAGCATCAGCAGCTGGCGGGTCAATTGSCAAAGGTGGAATAACCGTGAAACGAGTTGCA 1550
QY 1171 GAATTTGACGGCAGCTGAGGTGGTGTAGTACCAAGAGACAGACCCCTGATGAGAACACCA 1230
DB 1551 GAATTTGACGGCAGCTGAGGTGGTGTAGTACCAAGAGACAGACCCCTGATGAGAACACCA 1510
QY 1231 GGTTCATCGTGAATAATCATCGACATTTCTATGCCAGTCAAGTGGCTCAACCGGAAGATCCG 1290
DB 1611 GGTTCATCGTGAATAATCATCGACATTTCTATGCCAGTCAAGTGGCTCAACCGGAAGATCCG 1670
QY 1291 AGACATCTCGGCCAGGTTAAGCAGCAGCATCAGAGGGGACAGATAACCGGCCCGAGG 1350
DB 1671 AGACATCTCGGCCAGGTTAAGCAGCAGCATCAGAGGGGACAGATAACCGGCCCGAGG 1730
QY 1351 ACGGAGGAAGTACCGACGCCCTCCCTGTCCCTTNGAGTCCAGGACAAACACGGCGAGAA 1410

```

Db 1731 ACGGAGAGTACCAGCCCTCCCTGTCTCCCTTCCAGTCCAGGACCAACGGGAGAGA 1790
 QY 1411 TCAGAGTGTCTCTCCCGGAGGCTGAGAAATGAGTGGGAATCCGGGACACNTGGGCC 1470
 Db 1791 TCAGAGTGTCTCTCCCGGAGGCTGAGAAATGAGTGGGAATCCGGGACACCTGGGCC 1850
 QY 1471 GGGCTGTAGATCAGGTTTGCCACTTGTATTGAGAAAGATGTTCCAGTGAAGGAACCTGAT 1530
 Db 1851 GGGCTGTAGATCAGGTTTGCCACTTGTATTGAGAAAGATGTTCCAGTGAAGGAACCTGAT 1910
 QY 1531 CTNTRAGCCCAACACCCCAATATGGCCCAACACTGNTGCCCTCGGGGTGFCAGA 1590
 Db 1911 CTCTAGCCCAACACCCCAATATGGCCCAACACTGNTGCCCTCGGGGTGFCAGA 1970
 QY 1591 AATTNAGCCCAAGGACATTTTAAAGTGTATTGAGAAAGATGTTCCAGTGAAGGAACCTGAT 1650
 Db 1971 AATTCTAGCCCAAGGACATTTTAAAGTGTATTGAGAAAGATGTTCCAGTGAAGGAACCTGAT 2030
 QY 1651 AAGAGGTTGATCACACCTCAGTGGGAAGAAAATAAAATTTCTTTCAGGTTTAAAA 1708
 Db 2031 AAGAGGTTGATCACACCTCAGTGGGAAGAAAATAAAATTTCTTTCAGGTTTAAAA 2088

RESULT 8

AF198254 2381 bp mRNA linear PRI 02-MAR-2000
 LOCUS Homo sapiens mRNA-binding protein CRDBP mRNA, complete cds.
 DEFINITION AF198254
 ACCESSION AF198254
 VERSION AF198254.1 GI:7141071

KEYWORDS

ORGANISM

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

FEATURES

source

1..2381

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/chromosomes="17"

/map="17q21.3-q22"

/cell_line="MCF-7"

306..2039

/note="KH-domain containing protein; similar to Mus

musculus CRD-BP (c-myc mRNA coding region

determinant-binding protein) and Homo sapiens IMP-1

(insulin growth factor II mRNA-binding protein)"

/codon_start=1

/product="mRNA-binding protein CRDBP"

/protein_id="AAE37203.1"

/db_xref="GI:7141072"

/translation="MNKLXIGNNESVTPADLEKVFPAEHKISYQGFIVKSGYAFVDC

PEHWAKAETFSQVLEQGRLEIHSVPKQSRKIQIRNIPPIRWELVLSLLA

QYGVNCEQVNTGESAVVNTYSNREQTRQIMKLNGHLENHALKVSYPDQIA

QGPENRRGSGRSGRQVPAAGAPAKQOQVDIPLRLVPTQVYGAIIKGEKATI

RNITKQTSKIDVHRKENGAAEKALSVHSTPEGSSACKMILETHKEADTKTAD

VLKTLAHNFVGRGLGKEGNLKKVQDTEFKITISSLODLTLNPERITVKALE

NCRARQETMKVREAYENDVAASLQSHLTPLNLAAVGLFPASSAVPPPPSVTG

AAPYSFMQAPQEMQVFIIPAQAVGAIIGKKGQHIKQLSRFASASIKIAPETPSK

VRMVIITGPPEAQFQAQGRYIKLEENFFGPKVEVKLETHIRVPASAGRVIKGGK

ORIGIN

Query Match 48.1%; Score 821; DB 9; Length 2381;
 Best Local Similarity 99.3%; Pred. No. 0;
 Matches 1491; Conservative 0; Mismatches 5; Indels 6; Gaps 1;

QY 37 AGCCATCATGAACTGAATGGCCACCAAGTGTGGAGAACCAATGCCCTGAAGGTCTCTACAT 96
 Db 707 AGCCATCATGAACTGAATGGCCACCAAGTGTGGAGAACCAATGCCCTGAAGGTCTCTACAT 766
 QY 97 CCCGATGAGCAGATAGCACAGGACCTGAGAATGGGCGCCGAGGGGCTTTGGCTCTCG 156
 Db 767 CCCGATGAGCAGATAGCACAGGACCTGAGAATGGGCGCCGAGGGGCTTTGGCTCTCG 826
 QY 157 GGGTCAGCCCCCGCAGGGGCTCACTGTGGCAGCGGGGCCCCCAGCCAGCAGCAGCAAGT 216
 Db 827 GGGTCAGCCCCCGCAGGGGCTCACTGTGGCAGCGGGGCCCCCAGCCAGCAGCAGCAAGT 886
 QY 217 GGACATCCCCCTTCGGCTCCCTGGTCCGCCACCCAGTATGTGGTGCCATTATGGCAAGGA 276
 Db 887 GGACATCCCCCTTCGGCTCCCTGGTCCGCCACCCAGTATGTGGTGCCATTATGGCAAGGA 946
 QY 277 GGGGCCCCACCATCCGCAACATCACAAAACAGACCCAGTCCAAGATAGACGTGTCATAGAA 336
 Db 947 GGGGCCCCACCATCCGCAACATCACAAAACAGACCCAGTCCAAGATAGACGTGTCATAGAA 1006
 QY 337 GGAGAACGACGTGACGTGAAAAGCCATCAGTGTGCACCTCCACCCCTGAGGCGTCTC 396
 Db 1007 GGAGAACGACGTGACGTGAAAAGCCATCAGTGTGCACCTCCACCCCTGAGGCGTCTC 1066
 QY 397 CTCGCTGTGAAGATGATCTTTGAGATATTCATATAAGAGAGCTAAGGACACCAAAACGGC 456
 Db 1067 CTCGCTGTGAAGATGATCTTTGAGATATTCATATAAGAGAGCTAAGGACACCAAAACGGC 1126
 QY 457 TGACAGGTTCCCTCGAAGATCCCTGGCCATAATAACTTTGTAGGGCGTCTCATGGCAA 516
 Db 1127 TGACAGGTTCCCTCGAAGATCCCTGGCCATAATAACTTTGTAGGGCGTCTCATGGCAA 1186
 QY 517 GGAAGGCGGAACTCTGAAGAGGTAGACAGATACCGAGACAAAATACCATCTCTCTC 576
 Db 1187 GGAAGGCGGAACTCTGAAGAGGTAGACAGATACCGAGACAAAATACCATCTCTCTC 1246
 QY 577 GTTCAAGACCTTACCTTTTACAAACCTTGAGAGACCATCACTGTGAAGGGGGCCATCGA 636
 Db 1247 GTTCAAGACCTTACCTTTTACAAACCTTGAGAGACCATCACTGTGAAGGGGGCCATCGA 1306
 QY 637 GAATTGTTCAGGGCGGAGCAGGAAATATGAAGAAAGTTCCGAGGCGCTATGAGAATGA 696
 Db 1307 GAATTGTTCAGGGCGGAGCAGGAAATATGAAGAAAGTTCCGAGGCGCTATGAGAATGA 1366
 QY 697 TGTGGTGCATGAGC-----TCTCACCTGATCCCTGGCCTGAACCTGGCTGTGTAGG 750
 Db 1367 TGTGGTGCATGAGCCTGCACTCTACCTGACCCCTGGCTGAACCTGGCTGTGTAGG 1426
 QY 751 TCTTTTCCAGCTTCATCCAGCGAGTCCCGCCGCTCCGAGCAGCGTTACTGGGGCTGC 810
 Db 1427 TCTTTTCCAGCTTCATCCAGCGAGTCCCGCCGCTCCGAGCAGCGTTACTGGGGCTGC 1486
 QY 811 TCCCTATAGCTCTTTTATGACGGCTCCCGAGCAGAGATGGTGCAGGTGTTTATCCCGCG 870
 Db 1487 TCCCTATAGCTCTTTTATGACGGCTCCCGAGCAGAGATGGTGCAGGTGTTTATCCCGCG 1546
 QY 871 CCAGCGAGTGGGCGCCATCATCGCAAGAGAGGGGCGAGCACAATCAACAGCTCTCCCGGTT 930
 Db 1547 CCAGCGAGTGGGCGCCATCATCGCAAGAGAGGGGCGAGCACAATCAACAGCTCTCCCGGTT 1606
 QY 931 TGCAGCGCTCCATCAAGATTGCAACCCGGAACCTGACTCCAAAGTTCGTATGGT 990
 Db 1607 TGCAGCGCTCCATCAAGATTGCAACCCGGAACCTGACTCCAAAGTTCGTATGGT 1666
 QY 991 TATCATCTGGACCGCCAGAGGGCCCCAATTAAGGCTCAGGGAAGAAATCTATGCGAAACT 1050

TYNELQNLTAEEVVVPRDQTPDENQVIVKLIIGHFYASQMAQRKIRDILAQVKQOHOK
 GSNQAQARRK"

Db 1667 TATCATCTGGACCGCCAGAGCCCAATTCAGGCTCAGGGAAGATCTATGGCAACT 1726
Qy 1051 CAAAGGAGGAGAACTTTCTTGGTCCCAAGAGGAGGAGTGAAGCTGGAGACCAACATAGTGT 1110
Db 1727 CAAAGGAGGAGAACTTTCTTGGTCCCAAGAGGAGGAGTGAAGCTGGAGACCAACATAGTGT 1786
Qy 1111 GCAGATCATGAGCTTGGCCGGGTCAATTCGCAAGGTGAAAAAAGGTGAACGAGTTGCA 1170
Db 1787 GCAGATCATGAGCTTGGCCGGGTCAATTCGCAAGGTGAAAAAAGGTGAACGAGTTGCA 1846
Qy 1171 GAATTTGACGGCAGCTGAGGTGATGATCAAGAGACCAAGACCCCTGATGAGAAAGCA 1230
Db 1847 GAATTTGACGGCAGCTGAGGTGATGATCAAGAGACCAAGACCCCTGATGAGAAAGCA 1906
Qy 1231 GGTCTCGTGAATCATCGGATTTCTATGCGCTCAGATGGCTCAACGGAAGATCG 1290
Db 1907 GGTCTCGTGAATCATCGGATTTCTATGCGCTCAGATGGCTCAACGGAAGATCG 1966
Qy 1291 AGACATCTTGGCCAGGTTAAGCAGCAGCATCAGAGGAGCAGAGTAACAGGCCCGAGC 1350
Db 1967 AGACATCTTGGCCAGGTTAAGCAGCAGCATCAGAGGAGCAGAGTAACAGGCCCGAGC 2026
Qy 1351 ACGGAGAGTGAACAGCCCTCCCTGCTGCTTNGAGTCAGAGACCAACAGCGGAGAA 1410
Db 2027 ACGGAGAGTGAACAGCCCTCCCTGCTGCTTNGAGTCAGAGACCAACAGCGGAGAA 2086
Qy 1411 TCAGAGTGTCTCTCCCGCAGGCTGAGATGAGTGGGAATCCGGACACNTGGGCC 1470
Db 2087 TCAGAGTGTCTCTCCCGCAGGCTGAGATGAGTGGGAATCCGGACACNTGGGCC 2146
Qy 1471 GGGCTGTAGATCAGTTGGCCACTTGTATGAGAAAGATGTTCCAGTGAGGAACCTGTAT 1530
Db 2147 GGGCTGTAGATCAGTTGGCCACTTGTATGAGAAAGATGTTCCAGTGAGGAACCTGTAT 2206
Qy 1531 CT 1532
Db 2207 CT 2208

RESULT 9
BD127811
LOCUS 2780 bp DNA linear PAT 18-SEP-2002
DEFINITION Primer for synthesizing full-length cDNA and use thereof.
ACCESSION BD127811
VERSION BD127811.1 GI:23222756
KEYWORDS JP 2002017375-A/3242.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 2780)
Ota,T., Nishikawa,T., Isogai,T., Hayashi,K., Ishii,S., Kawai,Y.,
Wakamatsu,A., Sugiyama,T., Nagai,K., Kojima,S., Otsuki,T. and
Koga,H.
TITLE Primer for synthesizing full-length cDNA and use thereof
JOURNAL Patent: JP 2002017375-A 3242 22-JAN-2002;
COMMENT HELIX RESEARCH INSTITUTE
OS Homo sapiens (human)
FN JP 2002017375-A/3242
PD 22-JAN-2002
PF 07-JUL-2000 JP 2000253172
PI TOSHIO OTA, TETSUO NISHIKAWA, TAKAO ISOGAI, KOJI HAYASHI, SHIZUKO
ISHII,
PI YURI KAWAI, AI WAKAMATSU, TOMOYASU SUGIYAMA, KEIICHI NAGAI, PI
SHINICHI KOJIMA,
PI TETSUJI OTSUKI, HISASHI KOGA
C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/ PC
10,
PC C12P21/02, C12Q1/68/ C12P21/08, G06F17/30, C12N15/00, C12N5/00 CC
Primer for synthesizing full-length cDNA and use thereof FH Key
Location/Qualifiers

FEATURES
source
FT CDS Location/Qualifiers
1..2780
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 43.4% Score 741; DB 6; Length 2780;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 991; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
Qy 713 TCTCACCCTGATCCCTGGCCTGAACCTGGTGCTGTAGGTCTTTTCCAGCTTCATCCAGC 772
Db 421 TCTCACCCTGATCCCTGGCCTGAACCTGGTGCTGTAGGTCTTTTCCAGCTTCATCCAGC 480
Qy 773 GCAGTCCCGCGCTCCACGACGGTTACTGGGGCTGCTCCCTATAGCTCCTTTATGCAG 832
Db 481 GCAGTCCCGCGCTCCACGACGGTTACTGGGGCTGCTCCCTATAGCTCCTTTATGCAG 540
Qy 833 GCTCCCGCAGCAGAGATGCTGACGCTGTTTATCCCGCCCGCAGCAGTGGCGGCATCATC 892
Db 541 GCTCCCGCAGCAGAGATGCTGACGCTGTTTATCCCGCCCGCAGCAGTGGCGGCATCATC 600
Qy 893 GGCAGAGGGGAGCAGCATCAACAGCTCTCCGGTTTCCAGCGGCTCCATCAAGATT 952
Db 601 GGCAGAGGGGAGCAGCATCAACAGCTCTCCGGTTTCCAGCGGCTCCATCAAGATT 660
Qy 953 GCACACCCGAACACCTGACTCCAAAGTTCGTATGTTTATCATCTGACGCGCCAGAG 1012
Db 661 GCACACCCGAACACCTGACTCCAAAGTTCGTATGTTTATCATCTGACGCGCCAGAG 720
Qy 1013 GCCCAATTCAGGCTCAGGGAAGAACTATGCGAAACTATGCGAAACTCAAGAGAGAACTCTTTGGT 1072
Db 721 GCCCAATTCAGGCTCAGGGAAGAACTATGCGAAACTCAAGAGAGAACTCTTTGGT 780
Qy 1073 CCCAAGGAGGAAGTGAAGCTGGAGACCCACATACGTGTGCGACATCAGCAGCTGCCCGG 1132
Db 781 CCCAAGGAGGAAGTGAAGCTGGAGACCCACATACGTGTGCGACATCAGCAGCTGCCCGG 840
Qy 1133 GTCAATTCGCAAAAGGTGAAAAACGGTGAACAGTTCGAGAAATTTGACGGCAGCTGAGGTG 1192
Db 841 GTCAATTCGCAAAAGGTGAAAAACGGTGAACAGTTCGAGAAATTTGACGGCAGCTGAGGTG 900
Qy 1193 GTAGTACCAAGAGACCAAGACCCCTGATGAGAACGACAGCTGATCGTGAATAATCATCGGA 1252
Db 901 GTAGTACCAAGAGACCAAGACCCCTGATGAGAACGACAGCTGATCGTGAATAATCATCGGA 960
Qy 1253 CATTTCTATGCCAGTCAGATGGCTCAACGGAAGATCCGAGACATCTCTGCCCAAGTTAAG 1312
Db 961 CATTTCTATGCCAGTCAGATGGCTCAACGGAAGATCCGAGACATCTCTGCCCAAGTTAAG 1020
Qy 1313 CAGCAGATCAGAAAGGAGCAGATTAACAGCCCGCAGCAGCGAGGAGTGAACAGCCCT 1372
Db 1021 CAGCAGATCAGAAAGGAGCAGATTAACAGCCCGCAGCAGCGAGGAGTGAACAGCCCT 1080
Qy 1373 CCTGTCTCTTNGAGTCCAGGACAAACAAACGGGAGAGAAATCGAGAGTGTCTCTCCCGGG 1432
Db 1081 CCTGTCTCTTNGAGTCCAGGACAAACAAACGGGAGAGAAATCGAGAGTGTCTCTCCCGGG 1140
Qy 1433 AGGCCTGAGAAATGAGTGGGAATCCGGACACNTGCGCCGGCTGTAGATCAGGTTGGCC 1492
Db 1141 AGGCCTGAGAAATGAGTGGGAATCCGGACACNTGCGCCGGCTGTAGATCAGGTTGGCC 1200
Qy 1493 ACTTGTATGAGAAAGATGTTCCAGTGAAGAAACCTGATCTNTCAGCCCAACACCCACC 1552
Db 1201 ACTTGTATGAGAAAGATGTTCCAGTGAAGAAACCTGATCTCTCAGCCCAACACCCACC 1260
Qy 1553 CAATTGCCCAACACCTGTNTGCCCTCGGGGTGTCTCAGAAATTTAGCGAAGCACTTTT 1612
Db 1261 CAATTGCCCAACACCTGTNTGCCCTCGGGGTGTCTCAGAAATTTAGCGAAGCACTTTT 1320
Qy 1613 AAACGTGGATTGTTTAAAGAGAGCTCTCCAGGCCCAACCAAGAGGGTGGATCAGACCTCAG 1672


```

Db      1321  AACGTGGATTGTTTAAAGAGCTCTCCAGGCCCCACAGAGGGTGGATCATCACCTCAG 1380
      |||||||
Qy      1673  TGGGAAGAAAATAAAATTCCTTCAGGTTTAAAA 1708
      |||||||
Db      1381  TGGGAAGAAAATAAAATTCCTTCAGGTTTAAAA 1416
      |||||||

RESULT 10
AK074915      2780 bp      mRNA      linear      PRI 03-SEP-2002
LOCUS      Homo sapiens cDNA FLJ90434 fis, clone NT2RP3000789, highly similar
DEFINITION      to Homo sapiens IGF-II mRNA-binding protein 1 (IMP-1) mRNA.
ACCESSION      AK074915
VERSION      AK074915.1 GI:22760672
KEYWORDS      oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens
ORGANISM      Homo sapiens
REFERENCE      1
AUTHORS      Isogai,T., Ota,T., Nishikawa,T., Hayashi,K., Sugano,S., Ishii,S.,
      Sugiyama,T., Suzuki,Y., Nagai,K., Yamamoto,J., Wakamatsu,A., Nakamura,Y.,
      Kawai-Ho,Y., Saito,K., Ono,T., Okano,K., Yoshikawa,Y.,
      Kojima,S., Nagahara,K., Masuho,Y., Hattori,A., Okumura,K., Iwayanagi,T. and
      Actuska,S., Sasaki,N., Hattori,A., Okumura,K., Iwayanagi,T. and
      Ninomiya,K.
TITLE      NEDO human cDNA sequencing project
JOURNAL      Unpublished
REFERENCE      2 (bases 1 to 2780)
AUTHORS      Isogai,T. and Otsuki,T.
TITLE      Direct Submission
JOURNAL      Submitted (25-MAR-2002) Takao Isogai, Helix Research Institute,
      Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
      (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
COMMENT      NEDO human cDNA sequencing project supported by Ministry of
      Economy, Trade and Industry of Japan; cDNA full insert sequencing:
      Research Association for Biotechnology; cDNA library construction:
      Institute of Medical Science, University of Tokyo, Laboratory of
      Genome Structure, Human Genome Center; cDNA 5'- & 3'-end one pass
      sequencing and clone selection: Helix Research Institute (supported
      by Japan Key Technology Center etc.).

FEATURES
      source
      1..2780
      /organism="Homo sapiens"
      /mol_type="mRNA"
      /db_xref="taxon:9606"
      /clone="NT2RP3000789"
      /cell_line="NT2"
      /cell_type="teratocarcinoma"
      /clone_lib="NT2RP3"
      /notes="cloning vector: pME18FL3-mRNA from NT2 neuronal
      precursor cells after 2-weeks retinoic acid (RA)
      induction"

ORIGIN
Query Match      43.4%; Score 741; DB 9; Length 2780;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 991; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy      713  TCTCACCTGATCCCTGGCCTGAACCTGGCTGCTGTAGGCTTTTCCAGCTTCATCCAGC 772
Db      421  TCTCACCTGATCCCTGGCCTGAACCTGGCTGCTGTAGGCTTTTCCAGCTTCATCCAGC 480
Qy      773  GCAGTCCCGCGGCTCCACAGAGGTACTGGGGCTGCTCCCTATAGCTCCTTTATGCAG 832
Db      481  GCAGTCCCGCGGCTCCACAGAGGTACTGGGGCTGCTCCCTATAGCTCCTTTATGCAG 540
Qy      833  GCTCCCGAGCAGAGATGGTGCAGGTGTTTATCCCGCCAGCAGTGGCGGCATCATC 892
Db      541  GCTCCCGAGCAGAGATGGTGCAGGTGTTTATCCCGCCAGCAGTGGCGGCATCATC 600
Qy      893  GGCAAGAGGGGAGGAGCATCAAAACAGCTCTCCCGGTTTGGCCAGCGCCTCCATCAAGATT 952

```

```

Db      601  GGCAGAGAGGGGAGCAGCATCAACAGCTCTCCCGGTTTGCAGGCGCTCCATCAAGATT 660
Qy      953  GCACACCCGAAAACACCTGACTCCAAAGTTTCGTATGGTTATCATCATCTGGACCCAGAG 1012
Db      661  GCACACCCGAAAACACCTGACTCCAAAGTTTCGTATGGTTATCATCATCTGGACCCAGAG 720
Qy      1013  GCCCAATTCAAGGCTCAGGGGAAGATCTATGGCAAACTCAAGAGGAGGAACTTCTTTGGT 1072
Db      721  GCCCAATTCAAGGCTCAGGGGAAGATCTATGGCAAACTCAAGAGGAGGAACTTCTTTGGT 780
Qy      1073  CCCAAGGAGGAAAGTGAAGCTGGAGCCACATACGTGTGCCAGCATCAGCAGCTGGCGG 1132
Db      781  CCCAAGGAGGAAAGTGAAGCTGGAGCCACATACGTGTGCCAGCATCAGCAGCTGGCGG 840
Qy      1133  GTCATTGGCAAGGTTGGAACCGTGAACCGTTCAGAGTTGAGAGATTTGACGGCAGCTGAGGTG 1192
Db      841  GTCATTGGCAAGGTTGGAACCGTGAACCGTTCAGAGTTGAGAGATTTGACGGCAGCTGAGGTG 900
Qy      1193  GTAGTACCAAGAGACCAGACCCCTGATGAGAACACACAGGTCATCGTGAATAATCATCGGA 1252
Db      901  GTAGTACCAAGAGACCAGACCCCTGATGAGAACACACAGGTCATCGTGAATAATCATCGGA 960
Qy      1253  CATTTCTATGCGTCAAGTGGCTCAACCGGAAGATCCGAGACATCTGGCCCGGTTAAG 1312
Db      961  CATTTCTATGCGTCAAGTGGCTCAACCGGAAGATCCGAGACATCTGGCCCGGTTAAG 1020
Qy      1313  CAGCAGCATCAGAGGGAGCAGAGTAACCCAGGCCACGAGCAGGAGTAAGTACCAGCCCT 1372
Db      1021  CAGCAGCATCAGAGGGAGCAGAGTAACCCAGGCCACGAGCAGGAGTAAGTACCAGCCCT 1080
Qy      1373  CCCTGTCCCTTNGAGTCCAGGACAAACACCGGCGAGAAATCGAGAGTGTCTCTCCCGGC 1432
Db      1081  CCCTGTCCCTTNGAGTCCAGGACAAACACCGGCGAGAAATCGAGAGTGTCTCTCCCGGC 1140
Qy      1433  AGGCTGAGAGTGTGGGAAATCCGGGACACNTGGCGCGGCTGTAGATCAGGTTTGCC 1492
Db      1141  AGGCTGAGAGTGTGGGAAATCCGGGACACNTGGCGCGGCTGTAGATCAGGTTTGCC 1200
Qy      1493  ACTTGATTGAGAAAGATGTTCCAGTGAAGAACCTGTATTCAGCCGCCAACACCCACC 1552
Db      1201  ACTTGATTGAGAAAGATGTTCCAGTGAAGAACCTGTATTCAGCCGCCAACACCCACC 1260
Qy      1553  CAATTGGCCCAACACTGNTGCCCTCGGGGTGTGAGAAATNTAGCCAGGACACTTTT 1612
Db      1261  CAATTGGCCCAACACTGNTGCCCTCGGGGTGTGAGAAATNTAGCCAGGACACTTTT 1320
Qy      1613  AAACGTGGATTGTTTAAAGAGCTCTCCAGGCCCCACCAAGAGGTTGGATCACACCTCAG 1672
Db      1321  AAACGTGGATTGTTTAAAGAGCTCTCCAGGCCCCACCAAGAGGTTGGATCACACCTCAG 1380
Qy      1673  TGGGAAGAAAATAAAATTCCTTCAGGTTTAAAA 1708
Db      1381  TGGGAAGAAAATAAAATTCCTTCAGGTTTAAAA 1416

RESULT 11
BD124998
LOCUS      BD124998      833 bp      DNA      linear      PAT 18-SEP-2002
DEFINITION      Primer for synthesizing full-length cDNA and use thereof.
ACCESSION      BD124998
VERSION      BD124998.1 GI:23219943
KEYWORDS      JP 2002017375-A/429
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
REFERENCE      1 (bases 1 to 833)
AUTHORS      Ota,T., Nishikawa,T., Isogai,T., Hayashi,K., Ishii,S., Kawai,Y.,
      Wakamatsu,A., Sugiyama,T., Nagai,K., Kojima,S., Otsuki,T. and
      Koga,H.
TITLE      Primer for synthesizing full-length cDNA and use thereof
JOURNAL      Patent: JP 2002017375-A 429 22-JAN-2002;

```

<hr/>					
COMMENT					
OS	Homo sapiens (human)				
PN	JP 2002017375-A/429				
PD	22-JAN-2002				
PF	07-JUL-2000 JP 2000253172				
PI	TOSHIO OTA, TETSUO NISHIKAWA, TAKAO ISOGAI, KOJI HAYASHI, SHIZUKO ISHII,				
PI	YURI KAWAI, AI WAKAMATSU, TOMOYASU SUGIYAMA, KEIICHI NAGAI, PI SHINICHI KOJIMA,				
PI	TETSUJI OTSUKI, HISASHI KOGA				
PC					
	C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/ PC 10, C12P21/02, C12Q1/68//C12P21/08, G06F17/30, C12N15/00, C12N5/00 CC Primer for synthesizing full-length cDNA and use thereof FH Key				
FT	source	Location/Qualifiers			
FT	1..833	/organism='Homo sapiens (human)'			
FEATURES					
source	Location/Qualifiers				
	1..833	/organism='Homo sapiens'			
		/mol_type='genomic DNA'			
		/db_xref='taxon:9606'			
ORIGIN					
	Query Match	24.2%; Score 414; DB 6; Length 833;			
	Best Local Similarity	100.0%; Pred. No. 2.5e-213;			
	Matches 414; Conservative	0; Mismatches 0; Indels 0; Gaps 0			
Qy	299	ACAAACGAGACCAGCTCCAAGATAGAGTGTCATAGGAGGAGAACCGACGTGCAGCTGAA 358			
Dd	1	ACAAACGAGACCAGCTCCAAGATAGAGTGTCATAGGAGGAGAACCGACGTGCAGCTGAA 60			
Qy	359	AAGCCATCAGTGTGCATCTCACCCCTGAGGGCTGCTCCCTCCCTGTTAAGATGATCTTG 418			
Dd	61	AAGCCATCAGTGTGCATCTCACCCCTGAGGGCTGCTCCCTCCCTGTTAAGATGATCTTG 120			
Qy	419	GAGATTATGCATAAAGAGGCTTAAGGACACCAAAAACGGCTGACGAGGTTCCCTCGAAGATC 478			
Dd	121	GAGATTATGCATAAAGAGGCTTAAGGACACCAAAAACGGCTGACGAGGTTCCCTCGAAGATC 180			
Qy	479	CTGGGCCCATATAACTTTGTAGGGCGTCTCATTTGGCAAGGAGCGAACCTGAAGAAG 538			
Dd	181	CTGGGCCCATATAACTTTGTAGGGCGTCTCATTTGGCAAGGAGCGAACCTGAAGAAG 240			
Qy	539	GTAGAGCAAGATACCGCAGACAAAAATCACCATCTCCTGTTGCAAGACCTTACCCCTTAC 598			
Dd	241	GTAGAGCAAGATACCGCAGACAAAAATCACCATCTCCTGTTGCAAGACCTTACCCCTTAC 300			
Qy	599	AACCTGTAGAGGACCATCACTGTGAAGGGGCCATCGAGAAATTTGTCAGGGCCGAGCAG 658			
Dd	301	AACCTGTAGAGGACCATCACTGTGAAGGGGCCATCGAGAAATTTGTCAGGGCCGAGCAG 360			
Qy	659	GAATAATCAGAGAAGTTCCGGAGGCGCTATGAGAATGATGTGCTGCCATGAC 712			
Dd	361	GAATAATCAGAGAAGTTCCGGAGGCGCTATGAGAATGATGTGCTGCCATGAC 414			
<hr/>					
RESULT 12					
BD126684					
LOCUS	833 bp DNA linear PAT 18-SEP-2002				
DEFINITION	Primer for synthesizing full-length cDNA and use thereof.				
ACCESSION	BD126684				
VERSION	BD126684.1 GI:23221629				
KEYWORDS	JP 2002017375-A/2115.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 833) OTA.T., Nishikawa,T., Isogai,T.; Hayashi,K., Ishii,S., Kawai,Y., Wakamatsu,A., Sugiyama,T., Nagai,K., Kojima,S., Otsuki,T. and Koga,H.				

TITLE JOURNAL REFERENCE AUTHORS

Homo sapiens chromosome 17, clone CTD-2244F11

Unpublished

2 (bases 1 to 107848)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Boguslavskiy,L., Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Fero,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N., Jones,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R., Kamat,A., Karatas,A., Kells,C., Lacroque,K., Maclean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., McPeeters,R., Meldrim,J., Menus,L., Mihova,T., Menga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Direct Submission

Submitted (22-DEC-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

3 (bases 1 to 107848)

Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A., Cooke,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Fero,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C., McCarthy,M., Meldrim,J., Menus,L., Mihova,T., Menga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schuback,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J., Testaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Direct Submission

Submitted (06-OCT-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

4 (bases 1 to 107848)

Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A., Cooke,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Fero,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C., McCarthy,M., Meldrim,J., Menus,L., Mihova,T., Menga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schuback,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J., Testaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Submitted (05-NOV-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Nov 5, 2002 this sequence version replaced gi:23506732.

TITLE JOURNAL COMMENT

All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L23011
Center clone name: 2244_F11

Only the initial 107.85 kb of this clone are being submitted.
The remainder of the clone is overlapped by accession number
AC091133 [WIGR project L12028].

FEATURES

source

1..107848
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="17"
/map="17"
/clone="CTD-2244F11"
/clone_lib="CITD1 Human BAC"
1..160
/rpt_family="LTR19A"
784..860
/rpt_family="L2"
1215..1493
/rpt_family="AluSx"
1498..1559
/rpt_family="(CA)n"
1591..1887
/rpt_family="L2"
2336..2530
/rpt_family="MIR"
2635..2937
/rpt_family="AluSx"
3488..3516
/rpt_family="AT-rich"
3517..3732
/rpt_family="AluJb"
complement(3798..4269)
/rpt_family="LTR33"
complement(4320..4414)
/rpt_family="MLTJ1"
4448..4744
/rpt_family="AluSx"
complement(4745..4791)
/rpt_family="MLTJ2"
complement(4868..4961)
/rpt_family="MLTJ1"
5740..5914
/rpt_family="LTR44"
6101..6244
/rpt_family="AluJc"
7572..7962
/rpt_family="MER57-int"
complement(8024..8187)
/rpt_family="L1MA4"
8188..8484
/rpt_family="AluSg"
complement(8485..8605)
/rpt_family="L1MA4"
8607..8878
/rpt_family="L1MA4"
9844..9992
/rpt_family="MER57-int"
10087..10397
/rpt_family="AluJb"
complement(10417..10647)

```

repeat_region      /rpt_family="MER30"
11922..112037
/rpt_family="FLAM_C"
repeat_region      complement(12063..12263)
/rpt_family="AluJo"
repeat_region      13037..13140
/rpt_family="MER57-int"
repeat_region      complement(13679..13786)
13816..14271
/rpt_family="AluSg/x"
repeat_region      14272..14393
/rpt_family="LTR44"
repeat_region      /rpt_family="FLAM_A"
14394..14434
/rpt_family="(TG)n"
repeat_region      14442..14477
/rpt_family="LTR44"
repeat_region      14566..14803
/rpt_family="MIR"
repeat_region      15796..15877
/rpt_family="MIR"
repeat_region      15989..16046
/rpt_family="L3"
repeat_region      16215..16602
/rpt_family="L2"
repeat_region      complement(16724..17021)
/rpt_family="AluSx"
repeat_region      17048..17408
/rpt_family="MLT1H"
repeat_region      17383..17414
/rpt_family="MLT1H"
repeat_region      complement(17510..17815)
/rpt_family="AluJb"
repeat_region      18437..18716
/rpt_family="AluSc"
repeat_region      18731..19060
/rpt_family="(TA)n"
repeat_region      19100..19261
/rpt_family="(TA)n"
repeat_region      19299..19323
/rpt_family="(CA)n"
repeat_region      complement(19513..19698)
/rpt_family="AluJb"

Query Match      16.6%; Score 284; DB 9; Length 107849;
Best Local Similarity 100.0%; Pred. No. 1e-142;
Matches 284; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 36 GAGCCATCATGAGCTGAATGGCCACAGTTGGAGAACCATCCCTGAGGTCCTCTACA 95
Db 83857 GAGCCATCATGAGCTGAATGGCCACAGTTGGAGAACCATCCCTGAGGTCCTCTACA 83798

QY 96 TCCCGATGAGCAGATAGCAGGACCTGAGAAATGGGCGCCGAGGGGCTTTGGCTCTC 155
Db 83797 TCCCGATGAGCAGATAGCAGGACCTGAGAAATGGGCGCCGAGGGGCTTTGGCTCTC 83738

QY 156 GGGGTAGCCCCCGCAGGCTACCTGTGGAGCGGGGGCCCCCAGCCAGCAGCAGG 215
Db 83737 GGGGTAGCCCCCGCAGGCTACCTGTGGAGCGGGGGCCCCCAGCCAGCAGCAGG 83678

QY 216 TGGACATCCCCCTTCCGCTCCTGGTGCACCCAGTATGTGGGTGCCATTATTGCAAGG 275
Db 83677 TGGACATCCCCCTTCCGCTCCTGGTGCACCCAGTATGTGGGTGCCATTATTGCAAGG 83618

QY 276 AGGGGCGCCACCATCCGACATCACAAAACAGCCAGTCGCAAG 319
Db 83617 AGGGGCGCCACCATCCGACATCACAAAACAGCCAGTCGCAAG 83574

```

RESULT 14

AC104974

LOCUS

DEFINITION Homo sapiens chromosome 17 clone RP11-145L16 map 17, WORKING DRAFT

154604 bp

DNA linear

HTG 13-FEB-2002

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

SEQUENCE, 23 ordered pieces.

AC104974

AC104974.3 GI:18653683

HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

Birren,B., Linton,L., Nusbaum,C. and Lander,E.

1 (bases 1 to 154604)

Homo sapiens chromosome 17, clone RP11-145L16

Unpublished

2 (bases 1 to 154604)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhgalter,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S., Dodge,S., Fazo,S.,
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Kamat,A., Karatas,A., Kells,C., Lacroque,K.,
Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrum,J.,
Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupbach,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliou,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Submitted (22-DEC-2001) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

On Feb 13, 2002 this sequence version replaced gi:18464184.

All repeats were identified using RepeatMasker:

Smit, A.P.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence.submissions@genome.wi.mit.edu

----- Project Information

Center project name: L22871

Center clone name: 145_L_16

----- Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 148031 bases at least Q40

Consensus quality: 150405 bases at least Q30

Insert size: 146000; agarose-fp

Insert size: 152404; sum-of-contigs

Quality coverage: 9.3 in Q20 bases; agarose-fp

Quality coverage: 8.9 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 23 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.

* This sequence will be replaced

* by the finished sequence as soon as it is available and

* the accession number will be preserved.

* 1 1709: contig of 1709 bp in length

* 1710 1809: Gap of 100 bp

* 1810 1999: contig of 190 bp in length
* 2000 2099: gap of 100 bp
* 2100 3127: contig of 1028 bp in length
* 3128 3227: gap of 100 bp
* 3228 4561: contig of 1334 bp in length
* 4562 5811: gap of 100 bp
* 5812 5911: contig of 1150 bp in length
* 5912 7151: gap of 100 bp
* 7152 7251: contig of 1240 bp in length
* 7252 8241: gap of 100 bp
* 8242 8361: contig of 1010 bp in length
* 8362 9740: gap of 100 bp
* 9740 12105: contig of 1378 bp in length
* 12106 12205: gap of 100 bp
* 12206 35760: contig of 2266 bp in length
* 35761 38000: gap of 100 bp
* 38001 38100: contig of 2355 bp in length
* 38101 41369: gap of 100 bp
* 41370 41469: contig of 2140 bp in length
* 41470 44094: contig of 3269 bp in length
* 44095 44194: gap of 100 bp
* 44195 47400: contig of 2625 bp in length
* 47401 47500: gap of 100 bp
* 47501 53741: contig of 3206 bp in length
* 53742 53841: gap of 100 bp
* 53842 60833: contig of 6241 bp in length
* 60834 60933: gap of 100 bp
* 60934 72421: contig of 6992 bp in length
* 72422 72521: gap of 100 bp
* 72522 84887: contig of 11488 bp in length
* 84888 84987: contig of 12366 bp in length
* 84988 99600: gap of 100 bp
* 99601 99700: contig of 14613 bp in length
* 99701 115353: gap of 100 bp
* 115354 115433: contig of 15653 bp in length
* 115434 131255: gap of 100 bp
* 131256 131356: contig of 15802 bp in length
* 131357 152505: gap of 100 bp
* 152506 152606: contig of 21150 bp in length
* 152607 152606: gap of 100 bp
* 152606 154604: contig of 1999 bp in length.

FEATURES

source
1. 154604
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="17"
/map="17"
/clone="RP11-145116"
/clone_lib="RPC1-11 Human Male BAC"
1. 1709
/note="assembly_fragment"
clone_end:SP6
vector_side:left
misc_feature 1810..1999
/note="assembly_fragment"
misc_feature 2100..3127
/note="assembly_fragment"
misc_feature 3228..4561
/note="assembly_fragment"
misc_feature 4662..5811
/note="assembly_fragment"
misc_feature 5912..7151
/note="assembly_fragment"
misc_feature 7252..8261
/note="assembly_fragment"
misc_feature 8362..9739
/note="assembly_fragment"
misc_feature 9840..12105
/note="assembly_fragment"
misc_feature 12206..35760

/note="assembly_fragment"
35861..38000
/note="assembly_fragment"
38101..41369
/note="assembly_fragment"
41470..44094
/note="assembly_fragment"
44195..47400
/note="assembly_fragment"
47501..53741
/note="assembly_fragment"
53842..60833
/note="assembly_fragment"
60934..72421
/note="assembly_fragment"
72522..84887
/note="assembly_fragment"
84988..99600
/note="assembly_fragment"
99701..115353
/note="assembly_fragment"
115454..131255
/note="assembly_fragment"
131356..152505
/note="assembly_fragment"
152606..154604
/note="assembly_fragment"
clone_end:T7
vector_side:right"
ORIGIN
Query Match 16.6%; Score 284; DB 2; Length 154604;
Best Local Similarity 100.0%; Pred. No. 1e-142;
Matches 284; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 36 GAGCCATCATGAAGCTGAATGGCCACCACTGGAGAACCATGCCCTGAAGTCTCCTACA 95
Db 133545 GAGCCATCATGAAGCTGAATGGCCACCACTGGAGAACCATGCCCTGAAGTCTCCTACA 133604
QY 96 TCCCGATGACGATACACAGGACCTGAGATGGCGCGGCGGGGCTTGGCTCTC 155
Db 133605 TCCCGATGACGATACACAGGACCTGAGATGGCGCGGCGGGGCTTGGCTCTC 133664
QY 156 GGGGTACGCGCCCGCCAGGGCTCACCTGTGGCAGCGGGGGCCCCCAGCCAGCAGCAAG 215
Db 133665 GGGGTACGCGCCCGCCAGGGCTCACCTGTGGCAGCGGGGGCCCCCAGCCAGCAGCAAG 133724
QY 216 TGGACATCCCTTCGGCTCTCTGGTGGCCACCCAGTATGGGTGCCATTATGGCAAG 275
Db 133725 TGGACATCCCTTCGGCTCTCTGGTGGCCACCCAGTATGGGTGCCATTATGGCAAG 133784
QY 276 AGGGGGCCACCATCGCAACATCAAAAACAGACCCAGTCCAAG 319
Db 133785 AGGGGGCCACCATCGCAACATCAAAAACAGACCCAGTCCAAG 133828

RESULT 15

AC025556/c 159122 bp DNA linear HTG 01-SEP-2000
LOCUS Homo sapiens chromosome 17 clone RP11-501C14, WORKING DRAFT
DEFINITION SEQUENCE, 17 unordered pieces.
ACCESSION AC025556
VERSION AC025556.4 GI:9958273
KEYWORDS HTG; HTGS PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 159122)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 159122)

AUTHORS TITLE JOURNAL

Waterston, R.H.
Direct Submission
Submitted (10-MAR-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Sep 1, 2000 this sequence version replaced gi:8570253.

COMMENT

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H NH0501C14
----- Summary Statistics -----
Sequencing vector: M13; 96%
Chemistry: Dye-terminator Big Dye; 4%
Assembly: Phrap; version 0.990319
Consensus quality: 148686 bases at least Q40
Consensus quality: 152251 bases at least Q30
Consensus quality: 154277 bases at least Q20
Insert size: 181000; agarose-fp
Insert size: 157522; sum-of-contigs
Quality coverage: 3.86 in Q20 bases; agarose-fp
Quality coverage: 4.48 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 17 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 7635: contig of 7635 bp in length
* 7636 7735: gap of unknown length
* 7736 15480: contig of 7745 bp in length
* 15481 15580: gap of unknown length
* 15581 25089: contig of 9509 bp in length
* 25090 25189: gap of unknown length
* 25190 35086: contig of 9817 bp in length
* 35087 35106: gap of unknown length
* 35107 44509: contig of 9403 bp in length
* 44510 44609: gap of unknown length
* 44610 56601: contig of 11992 bp in length
* 56602 56701: gap of unknown length
* 56702 64536: contig of 7835 bp in length
* 64537 64636: gap of unknown length
* 64637 79941: contig of 15305 bp in length
* 79942 80041: gap of unknown length
* 80042 95747: contig of 15705 bp in length
* 95748 95846: gap of unknown length
* 95847 125532: contig of 29686 bp in length
* 125533 126811: gap of unknown length
* 126812 126911: gap of unknown length
* 126912 129146: contig of 2235 bp in length
* 129147 129246: gap of unknown length
* 129247 132693: contig of 3446 bp in length
* 132694 132792: gap of unknown length
* 132793 135914: contig of 3122 bp in length
* 135915 136014: gap of unknown length
* 136015 143560: contig of 7546 bp in length
* 143561 143660: gap of unknown length
* 143661 150767: contig of 7106 bp in length
* 150768 150866: gap of unknown length
* 150867 159122: contig of 8256 bp in length.

FEATURES source

1..159122
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

/chromosome="17"
/clone="RP11-501C14"
1..7635
/note="assembly_name:Contig10"
7736..15480
/note="assembly_name:Contig11"
15581..25089
/note="assembly_name:Contig12"
25190..35006
/note="assembly_name:Contig13"
35107..44509
/note="assembly_name:Contig14"
44610..56601
/note="assembly_name:Contig15"
56702..64536
/note="assembly_name:Contig16"
64637..79941
/note="assembly_name:Contig17"
80042..95746
/note="assembly_name:Contig18"
95847..125532
/note="assembly_name:Contig19"
125633..126811
/note="assembly_name:Contig3"
126912..129146
/note="assembly_name:Contig4"
129247..132692
/note="assembly_name:Contig6"
132793..135914
/note="assembly_name:Contig5"
136015..143560
/note="assembly_name:Contig7"
143661..150766
/note="assembly_name:Contig8"
150867..159122
/note="assembly_name:Contig9"

ORIGIN

Query Match 16.6%; Score 284; DB 2; Length 159122;
Best Local Similarity 100.0%; Pred. No. 1e-142;
Matches 284; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 36 GAGCCATCATGAAGCTGAATGGCCACCAGTTGGAGAACCATGCCCTGAAGGTCTCTCTACA 95
Db 67231 GAGCCATCATGAAGCTGAATGGCCACCAGTTGGAGAACCATGCCCTGAAGGTCTCTCTACA 67172
QY 96 TCCCGGATGACAGATAGCAGGACCTGAGATGGGCGCCGAGGGGGCTTTGGCTCTC 155
Db 67171 TCCCGGATGACAGATAGCAGGACCTGAGATGGGCGCCGAGGGGGCTTTGGCTCTC 67112
QY 156 GGGGTACGCCCCCGCCAGGGCTCACCTGTGCGAGCGGGGGCCCGAGCCAGCAGCAAG 215
Db 67111 GGGGTACGCCCCCGCCAGGGCTCACCTGTGCGAGCGGGGGCCCGAGCCAGCAGCAAG 67052
QY 216 TGGACATCCCGCTTCGGCTCCTGGTCCCAACCATGATGTGGGTGCAATTTGGCAAGG 275
Db 67051 TGGACATCCCGCTTCGGCTCCTGGTCCCAACCATGATGTGGGTGCAATTTGGCAAGG 66992
QY 276 AGGGGGCCACATCCCGCAACATCACAAAACAGACCCAGTCCCAAG 319
Db 66991 AGGGGGCCACATCCCGCAACATCACAAAACAGACCCAGTCCCAAG 66948

Search completed: July 14, 2004, 01:45:04
Job time : 6733 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: July 13, 2004, 12:13:53 ; Search time 81.5 Seconds
(without alignments)
11842.727 Million cell updates/sec

Title: US-09-270-437D-5

Perfect score: 560

Sequence: 1 agggacgctgcgcaccgcc.....atttccttcagggttttaaaa 1708

Scoring table:

Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1586107 seqs, 282547505 residues

Word size: 1

Total number of hits satisfying chosen parameters: 2987630

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp
-Q=/cgn2.1/USPTO.spool_p/US09270437/runat.13072004.121955.9961/app.query.fasta.1.1863
-DB=A_Geneseq_29Jan04 -QFM=fastan -SUFFIX=rag -MINMATCH=0.1 -LOOPGL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09270437@cgn.1.1.154@runat.13072004.121955.9961 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

A_Geneseq_29Jan04:*
1: Geneseq1980s:*
2: Geneseq1990s:*
3: Geneseq2000s:*
4: Geneseq2001s:*
5: Geneseq2002s:*
6: Geneseq2003as:*
7: Geneseq2003bs:*
8: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	216	38.6	261	AAU16161	AAU16161 Human nov
2	216	38.6	261	ABU55230	ABU55230 Human nov
3	216	38.6	319	AAU93826	AAU93826 Human pol
4	166	29.6	577	AAU30649	AAU30649 A murine
5	116	20.7	250	AAU16579	AAU16579 Human nov
6	116	20.7	250	ABU55648	ABU55648 Human nov
7	36	6.4	579	ABU11328	ABU11328 Human nov
8	36	6.4	579	ABU11365	ABU11365 Human nov
9	36	6.4	579	ABU74960	ABU74960 Human nov
10	36	6.4	579	ABU75053	ABU75053 Human nov

11	36	6.4	579	5	ABB74997	Human lun
12	36	6.4	579	5	ABB75054	Human lun
13	36	6.4	579	5	ABP61917	Human lun
14	36	6.4	579	5	ABP61974	Human lun
15	36	6.4	579	5	ABP61880	Human lun
16	36	6.4	579	5	ABP61973	Human lun
17	36	6.4	579	7	ADA28536	Recombina
18	36	6.4	579	7	ADA28539	Recombina
19	36	6.4	579	7	ADA28438	Human lun
20	36	6.4	579	7	ADA28266	Human lun
21	36	6.4	579	7	ADA28266	Human lun
22	36	6.4	579	7	ADA28266	Human lun
23	36	6.4	583	4	ABG12592	Novel hum
24	36	6.4	586	5	ABP61968	Human lun
25	36	6.4	586	5	ABP61968	Human lun
26	36	6.4	586	7	ADA28517	Recombina
27	36	6.4	619	4	ABG211963	Novel hum
28	24	4.3	148	4	ABG211962	Novel hum
29	21	3.8	93	4	AAU38501	Peptide #
30	21	3.8	93	4	AAU78238	Human don
31	20	3.6	20	5	ABP61962	Human lun
32	20	3.6	20	5	ABP61961	Human lun
33	20	3.6	20	5	ABP61961	Human lun
34	20	3.6	20	5	ABP61962	Human lun
35	20	3.6	20	7	ADA28504	Human lun
36	20	3.6	20	7	ADA28505	Human lun
37	19	3.4	171	4	AAU16166	Human nov
38	19	3.4	171	4	AAU16583	Human nov
39	19	3.4	171	6	ABU55235	Human nov
40	19	3.4	171	6	ABU55652	Human nov
41	19	3.4	556	5	ABG96346	Human ova
42	19	3.4	594	4	ABG06795	Novel hum
43	19	3.4	614	4	ABG06794	Novel hum
44	19	3.4	620	4	AAU16163	Human nov
45	19	3.4	620	6	ABU55232	Human nov

ALIGNMENTS

RESULT 1

AAU16161

ID AAU16161 standard; protein; 261 AA.

XX AAU16161;

XX 07-NOV-2001 (first entry)

XX Human novel secreted protein, Seq ID 1114.

Human: immunosuppressive; antiarthritic; antirheumatic; cytostatic;
cardiant; vasotropic; cerebroprotective; neurotropic; neuroprotective;
antibacterial; virucide; fungicide; ophthalmological; vulnary;
secreted protein; rheumatoid arthritis; hyperproliferative disorder;
cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
cerebral ischaemia; angiogenesis; nervous system disorder;
Alzheimer's disease; infection; ocular disorder; corneal infection;
wound healing; epithelial cell proliferation; skin ageing; food additive;
preservative; antiproliferative.

OS Homo sapiens.

XX WO200155322-A2.

XX 02-AUG-2001.

XX 17-JAN-2001; 2001WO-US001341.

XX 31-JAN-2000; 2000US-0179065P.

XX 04-FEB-2000; 2000US-0180628P.

XX 24-FEB-2000; 2000US-0184664P.

XX 02-MAR-2000; 2000US-0186350P.

XX 16-MAR-2000; 2000US-0198874P.

XX 17-MAR-2000; 2000US-0190076P.

CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
 CC rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
 CC in diagnosing a pathological condition or the susceptibility to a
 CC pathological condition. Antibodies to the proteins can also be used in
 CC alleviating symptoms associated with the disorders and in diagnostic
 CC immunoassays e.g. radioimmunoassays or enzyme linked immunosorbent assays
 CC (ELISA). Disorders which are diagnosed or treated include autoimmune
 CC diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.
 CC neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac
 CC arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis,
 CC nervous system disorders e.g. Alzheimer's disease, infections caused by
 CC bacteria, viruses and fungi and ocular disorders e.g. corneal infection,
 CC and many other disorders listed in the specification. The polypeptides
 CC can also be used to aid wound healing and epithelial cell proliferation,
 CC to prevent skin aging due to sunburn, to maintain organs before
 CC transplantation, for supporting cell culture of primary tissues, to
 CC regenerate tissues and in chemotaxis. The polypeptides can also be used
 CC as a food additive or preservative to increase or decrease storage
 CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,
 CC minerals, cofactors and other nutritional components. The present
 CC sequence represents a novel secreted protein of the invention. Note: The
 CC sequence data for this patent did not form part of the printed

Alignment Scores:

Pred. No.: 6.67e-204 Length: 261
 Score: 216.00 Matches: 216
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 38.57% Indels: 0
 DB: 4 Gaps: 0

US-09-270-437D-5 (1-1708) x AAU16161 (1-261)

QY 713 TCTCAGCTGATCCCTGGCTGAACTGGTGTGTAGGTCTTTTCCAGCTTCATCCAGC 772
 Db 46 SerHisLeuIleProGlyLeuAsnLeuAlaValGlyLeuPheProAlaSerSerSer 65
 QY 773 GCAGTCCGCGGCTCCAGCAGCGTTACTGGGGCTGCTCCCTATAGCTCCTTTATGCGAG 832
 Db 66 AlaValProProProSerSerValThrGlyAlaAlaProTyrSerSerPheMetGln 85
 QY 833 GCTCCCGAGCAGGAGATGGTCAGGTGTTTATCCCGCCAGCAGCTGGCGGCATCATC 892
 Db 86 AlaProGluGlnMetValGlnValPheIleProAlaGlnAlaValGlyAlaIleIle 105
 QY 893 GGCAAGAGGGGAGCAGCATCAACAGCTCTCCCGGTTTGCCAGCGCCCTCCATCAAGATT 952
 Db 106 GlyLysLysGlyGlnHisIleLysGlnLeuSerArgPheAlaSerAlaSerIleLysIle 125
 QY 953 GCACCCAGGAAACCTGACTCCAAAGTTCGATGTTATCATCTGACCTGACCGCCAGAG 1012
 Db 126 AlaProProGluThrProAspSerLysValArgMetValIleIleThrGlyProProGlu 145
 QY 1013 GCCCAATTCAAGGCTCAGGGAAGAATCTATGGCAAACTCAAGGAGGAGAACTTCTTTGGT 1072
 Db 146 AlaGlnPheLysAlaGlnGlyArgIleTyrGlyLysLeuLysGluLysGluAsnPhePheGly 165
 QY 1073 CCAAGAGGAAGTGAAGCTGGAGACCCATACGTGTCCAGCATCAGCAGCTGGCCCGG 1132
 Db 166 ProLysGluGluValLysLeuGluThrHisIleArgValProAlaSerAlaAlaGlyArg 185
 QY 1133 GTCAATTGGCAAGGTGGAAAAACGGTCAACAGCTTCAGAAATTTGACGGCAGCTGAGGTG 1192
 Db 186 ValIleGlyGlyGlyLysThrValAsnGluLeuGlnAsnLeuThrAlaAlaGluVal 205
 QY 1193 GTAGTACCAAGAGCAGACCCCTGATGAGAACAGCAGCATCTGCTGAAATCATCGGA 1252
 Db 206 ValValProArgAspGlnThrProAspGluAsnAspGlnValIleValIleGly 225
 QY 1253 CATTTCTATCCAGTCAGTGGCTCAACGAGATCCGAGACATCTGTCGCCCGCTTATG 1312
 Db 226 HisPheTyrAlaSerGlnMetAlaGlnArgLysIleArgAspIleLeuAlaGlnValLys 245

QY 1313 CAGCAGCATCAGAGGACAGACAGTAACCCAGCCCGCAGCGAGGAG 1360
 Db 246 GlnGlnHisGlnLysGlyGlnSerAsnGlnAlaGlnAlaArgArgLys 261
 RESULT 2
 ABUS5230
 ID ABUS5230 standard; protein; 261 AA.
 XX AC ABUS5230;
 DT 18-MAR-2003 (first entry)
 DE Human novel polypeptide #317.
 KW Human; neural disorder; immune system disorder; renal disorder;
 KW muscular disorder; respiratory disease; reproductive disorder;
 KW gastrointestinal disorder; pulmonary disorder; cardiovascular disorder;
 KW hyperproliferative disorder; inflammatory disease; allergic reaction;
 KW blood related disorder; cancer; immunosuppressive; antiinflammatory;
 KW cardiovascular; nephrotropic; cytostatic; antiallergic; thrombolytic;
 KW haemostatic; antiarteriosclerotic.
 XX OS Homo sapiens.
 XX PN US2002132753-A1.
 XX PD 19-SEP-2002.
 XX PF 17-JAN-2001; 2001US-00764864.
 XX PR 31-JAN-2000; 2000US-0179085P.
 PR 04-FEB-2000; 2000US-0180628P.
 PR 28-JUN-2000; 2000US-0214886P.
 PR 07-JUL-2000; 2000US-0216647P.
 PR 07-JUL-2000; 2000US-0216880P.
 PR 11-JUL-2000; 2000US-0217487P.
 PR 11-JUL-2000; 2000US-0217486P.
 PR 14-JUL-2000; 2000US-0218290P.
 PR 26-JUL-2000; 2000US-0220963P.
 PR 14-AUG-2000; 2000US-0220964P.
 PR 14-AUG-2000; 2000US-0224518P.
 PR 14-AUG-2000; 2000US-0224519P.
 PR 14-AUG-2000; 2000US-0225267P.
 PR 14-AUG-2000; 2000US-0225268P.
 PR 14-AUG-2000; 2000US-0225270P.
 PR 14-AUG-2000; 2000US-0225447P.
 PR 14-AUG-2000; 2000US-0225757P.
 PR 14-AUG-2000; 2000US-0225758P.
 PR 22-AUG-2000; 2000US-0226868P.
 PR 30-AUG-2000; 2000US-0228924P.
 PR 01-SEP-2000; 2000US-0229287P.
 PR 01-SEP-2000; 2000US-0229343P.
 PR 01-SEP-2000; 2000US-0229344P.
 PR 01-SEP-2000; 2000US-0229345P.
 PR 05-SEP-2000; 2000US-0229509P.
 PR 05-SEP-2000; 2000US-0229513P.
 PR 08-SEP-2000; 2000US-0231413P.
 PR 21-SEP-2000; 2000US-0234223P.
 PR 21-SEP-2000; 2000US-0234274P.
 PR 25-SEP-2000; 2000US-0234997P.
 PR 27-SEP-2000; 2000US-0235834P.
 PR 29-SEP-2000; 2000US-0236327P.
 PR 29-SEP-2000; 2000US-0236367P.
 PR 29-SEP-2000; 2000US-0236368P.
 PR 29-SEP-2000; 2000US-0236369P.
 PR 29-SEP-2000; 2000US-0236370P.
 PR 02-OCT-2000; 2000US-0236802P.
 PR 02-OCT-2000; 2000US-0237037P.
 PR 02-OCT-2000; 2000US-0237038P.
 PR 02-OCT-2000; 2000US-0237039P.
 PR 02-OCT-2000; 2000US-0237040P.
 PR 13-OCT-2000; 2000US-0239935P.
 PR 20-OCT-2000; 2000US-0240960P.

PR 20-OCT-2000; 2000US-02411785P.
 PR 20-OCT-2000; 2000US-02411809P.
 PR 01-NOV-2000; 2000US-0244617P.
 PR 17-NOV-2000; 2000US-0249299P.
 PR 08-DEC-2000; 2000US-02511856P.
 PR 08-DEC-2000; 2000US-02511868P.
 PR 08-DEC-2000; 2000US-02511869P.
 XX (ROSE/) ROSEN C A.
 PA (RUBE/) RUBEN S M.
 PA (BARA/) BARASH S C.
 XX
 PI Rosen CA, Ruben SM, Barash SC;
 XX
 XX WPI; 2003-147444/14.
 DR N-PSDB; ABX73489.
 XX
 XX New polypeptides and nucleic acids, useful in gene therapy for treating,
 PT inhibiting or preventing e.g. neural, immune system, muscular,
 PT respiratory, reproductive, gastrointestinal, pulmonary, cardiovascular or
 PT renal disorders.
 XX
 PS Claim 11; SEQ ID NO 1114; 402pp; English.
 XX
 XX The invention relates to human novel polypeptides and their associated
 CC polynucleotides. The polypeptides and polynucleotides are useful in gene
 CC therapy for treating, inhibiting or preventing neural disorders, immune
 CC system disorders (e.g. systemic lupus erythematosus, rheumatoid arthritis
 CC and multiple sclerosis), muscular disorders, respiratory diseases (e.g.
 CC nasal vestibulitis, nasal polyps and sinusitis), reproductive disorders,
 CC gastrointestinal disorders, pulmonary disorders, cardiovascular disorders
 CC (e.g. congenital heart defects, Ebstein's anomaly and hypoplastic left
 CC heart syndrome), renal disorders (e.g. acute kidney failure and end-stage
 CC renal disease), hyperproliferative disorders (e.g. Hodgkin's disease and
 CC leukaemia), inflammatory diseases (e.g. septic shock, bursitis and
 CC appendicitis), allergic reactions and conditions (e.g. asthma), blood
 CC related disorders (e.g. thrombosis, atherosclerosis and myocardial
 CC infarction) and cancerous diseases. Sequences ABU54914-ABU55699 and
 CC ABU55748 represent human novel polypeptides of the invention
 XX
 SQ Sequence 261 AA;
 Alignment Scores:
 Pred. No.: 6.67e-204 Length: 261
 Score: 216.00 Matches: 216
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 38.57% Indels: 0
 DB: 6 Gaps: 0
 US-09-270-437D-5 (1-1708) x ABU55230 (1-261)
 QY 713 TCTCAGTCATCCCTGGCTGAACCTGGCTGTGTAGGTCTTTTCCAGCTTCATCCAGC 772
 Db 46 SerHisLeuIleProGlyLeuAsnLeuAlaAlaValGlyLeuPheProAlaSerSer 65
 QY 773 GCAGTCCCGCGCTCCAGCAGCTTACTGGGCTGCTCCCTATAGCTCTTTATGAC 832
 Db 66 AlaValProProProSerSerValThrGlyAlaAlaProTySerSerPheMetGln 85
 QY 833 GCTCCCGAGCAGGAGTGTGTCAGGTGTTATCCCGCCGAGCGATGGCGCCCATCATC 892
 Db 86 AlaProGluGlnGluMetValGlnValPheIleProAlaGlnAlaValGlyAlaIleIle 105
 QY 893 GGCAAGAGGGGAGCAGCATCAACAGCTCTCCCGGTTTCCAGCGCTCCATCAAGATT 952
 Db 106 GlyLysGlyGlnHisIleLysGlnLeuSerArgPheAlaSerAlaSerIleLysIle 125
 QY 953 GCACCAACCCGAAACACTGACTCCAAAGTTTCGTATGTTATCATCACTGCAGCCGACAG 1012
 Db 126 AlaProProGluThrProAspSerLysValArgMetValIleIleThrGlyProProGlu 145
 QY 1013 GCCCAATTCAAGCTCAGGGAAGAATCTATGGCAAACTCAAGAGGAGAACTTCTTTGGT 1072

Db 146 AlaGlnPheLysAlaGlnGlyArgIleTyGlyLysLeuLysGluGluAsnPhePheGly 165
 QY 1073 CCCAAGAGGAAGTGAAGCTGGAGACCCACATACGTGTGCAGCATCAGCAGCTGGCCGG 1132
 Db 166 ProLysGluGluValLysLeuGluThrHisIleArgValProAlaSerAlaAlaGlyArg 185
 QY 1133 GTCAATGGCAAGGTGCAAAAACGGTGAACGAGTTCAGAAATTTGACGGCAGCTGAGGTG 1192
 Db 186 ValIleGlyLysGlyGlyLysThrValAsnGluLeuGlnAsnLeuThrAlaAlaGluVal 205
 QY 1193 GTAGTACCAAGAGACACAGACCCCTGTAGAGAACGACACAGCTCATCGTGAATCATCGGA 1252
 Db 206 ValValProArgAspGlnThrProAspGluAsnAspGlnValIleValIleIleGly 225
 QY 1253 CATTCTATCCAGTCAGATGGCTCAACGGAAGATCCGAGACATCTGSCCAGGTTAAG 1312
 Db 226 HisPheTyAlaSerGlnMetAlaGlnArgLysIleArgAspIleLeuAlaGlnValLys 245
 QY 1313 CAGCAGCATCAGAAAGGACAGAGTAACAGGCCCGCCAGCAGCAGGAGGAG 1360
 Db 246 GlnGlnHisGlnLysGlyGlnSerAsnGlnAlaGlnAlaArgArgLys 261
 RESULT 3
 ID AAM93826
 ID AAM93826 standard; protein; 319 AA.
 AC AAM93826;
 XX
 DT 06-NOV-2001 (first entry)
 XX
 DE Human polypeptide, SEQ ID NO: 3887.
 XX
 KW Human; full length cDNA; cDNA synthesis; oligo-capping.
 XX
 CS Homo sapiens.
 XX
 PN EP1130094-A2.
 XX
 PD 05-SEP-2001.
 XX
 PF 07-JUL-2000; 2000EP-00114089.
 XX
 PR 08-JUL-1999; 99JP-00194486.
 PR 11-JAN-2000; 2000JP-00118774.
 PR 02-MAY-2000; 2000JP-00183765.
 PA (HELI-) HELIX RES INST.
 XX
 PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
 PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
 DR WPI; 2001-524255/58.
 DR N-PSDB; AAK94782.
 XX
 PT 830 Primers useful for synthesizing full length cDNA clones and their use
 in genetic manipulation.
 PS Claim 8; SEQ ID NO 3887; 1380pp + Sequence Listing; English.
 XX
 XX The invention relates to primers for synthesising full length cDNA
 CC clones. 830 cDNA molecules encoding a human protein have been isolated
 CC and nucleotide sequences of 5'- and 3'-ends of the cDNA molecules have
 CC been determined. Primers for synthesising the full length cDNA are useful
 CC for clarifying the function of the protein encoded by the cDNA. The full
 CC length clones were obtained by construction of full length enriched cDNA
 CC libraries that were synthesised by the oligo-capping method. The primers
 CC enable the production of the full length cDNA easily without any special
 CC methods. The present sequence is a polypeptide encoded by a full length
 CC human cDNA of the invention. Note: The sequence data for this patent did
 CC not form part of the printed specification, but was obtained in CD-ROM
 CC format directly from EPO
 XX

XX	PD	XX	16-SEP-1999.
XX	XX	XX	05-MAR-1999; 99WO-US004897.
XX	PF	XX	09-MAR-1998; 98US-0077372P.
XX	PR	XX	(WISC) WISCONSIN ALUMNI RES FOUND.
XX	PA	XX	Ross J;
XX	PI	XX	WPI; 1999-551506/46.
XX	DR	XX	N-P8DB; AA210617.
XX	DR	XX	Diagnosing presence or absence of a tumor in a human by examining c-myc coding region determinant-binding protein.
XX	PT	XX	Example; Fig 1A-D; 79pp; English.
XX	PS	XX	The present sequence represents a murine c-myc coding region determinant binding protein (CRD-BP). The presence or absence of a tumor can be determined by determining the levels of CRD-BP present in the suspect tissue, where the CRD-BP shields c-myc RNA from endonucleolytic attack and so prolongs its half-life. The methods are used for diagnosing presence or absence of a tumor in a human, especially breast, colon and pancreatic cancer. They are also used to inhibit cancer cell growth
XX	CC	XX	Sequence 577 AA;
XX	CC	XX	Alignment Scores:
XX	CC	XX	Pred. No.: 1,45e-154 Length: 577
XX	CC	XX	Score: 166.00 Matches: 166
XX	CC	XX	Percent Similarity: 100.00% Conservatative: 0
XX	CC	XX	Best Local Similarity: 100.00% Mismatches: 0
XX	CC	XX	Query Match: 29.64% Indels: 0
XX	CC	XX	DB: 2 Gaps: 0
XX	CC	XX	US-09-270-437D-5 (1-1708) x AAY30649 (1-577)
XX	CC	XX	QY 215 GTGGACATCCCTTCGGCTCGTCCGCCACCCAGTATGTGGTGCCATTATGGCAAG 274
XX	CC	XX	Db 194 ValAspIleProLeuArgLeuValProThrGlnTyrValGlyAlaIleGlyLys 213
XX	CC	XX	QY 275 GAGGGGGCCACATCCGCAACATCACAACACAGACCCAGTCCAAAGATAGACGTGCATAGG 334
XX	CC	XX	Db 214 GluGlyAlaThrIleArgAsnIleThrLysGlnThrGlnSerLysIleAspValHisArg 233
XX	CC	XX	QY 335 AAGGAGAACGGAGTGCAGCTGAGTGAAGAACCATCATGTGTGCACCTCCACCCCTGAGGGCTGC 394
XX	CC	XX	Db 234 LysGluAsnAlaGlyAlaAlaGluLysAlaIleSerValHisSerThrProGluGlyCys 253
XX	CC	XX	QY 395 TCCTCCGCTTCTAGATGATCTTCGAGATTATGCATAAAGAGCTAAGGACCAACAAACG 454
XX	CC	XX	Db 254 SerSerAlaCysLysMetIleGluGluIleValHisLysGluAlaLysAspThrLysThr 273
XX	CC	XX	QY 455 GGTGACGAGGTTCCTTCGAGATCCTCGCCCAATAAATCTTGTAGGGCGCTCTCATGGC 514
XX	CC	XX	Db 274 AlaAspGluValProLeuLysIleLeuAlaHisAsnAsnPheValGlyArgLeuIleGly 293
XX	CC	XX	QY 515 AAGGAGGACGGAACCTTGAGAGAGTACGAGACCATCATCTGTGAAGGGGGGCATC 634
XX	CC	XX	Db 294 LysGluGlyArgAsnLeuLysLysValGluGlnAspThrGluThrLysIleThrIleSer 313
XX	CC	XX	QY 575 TCGTTGCAAGACCTTACCTTTTACACCCCTGAGAGACCATCATCTGTGAAGGGGGGCATC 634
XX	CC	XX	Db 314 SerLeuGlnAspLeuThrLeuTyrAsnProGluArgThrIleThrValLysGlyAlaIle 333
XX	CC	XX	QY 635 GAGAAATCTTCGAGGGCGGAGCAGGAAATATGAGAAAGTTCGGAGGCGCTATGAGAT 694
XX	CC	XX	Db 334 GluAsnCysCysArgAlaGluGlnGluIleMetLysLysValArgGluAlaTyrGluAsn 353
XX	CC	XX	QY 695 GATGTGGCTGCCATGAGC 712
XX	CC	XX	US-09-270-437D-5 (1-1708) x AAY30649 (1-577)
XX	CC	XX	QY 215 GTGGACATCCCTTCGGCTCGTCCGCCACCCAGTATGTGGTGCCATTATGGCAAG 274
XX	CC	XX	Db 194 ValAspIleProLeuArgLeuValProThrGlnTyrValGlyAlaIleGlyLys 213
XX	CC	XX	QY 275 GAGGGGGCCACATCCGCAACATCACAACACAGACCCAGTCCAAAGATAGACGTGCATAGG 334
XX	CC	XX	Db 214 GluGlyAlaThrIleArgAsnIleThrLysGlnThrGlnSerLysIleAspValHisArg 233
XX	CC	XX	QY 335 AAGGAGAACGGAGTGCAGCTGAGTGAAGAACCATCATGTGTGCACCTCCACCCCTGAGGGCTGC 394
XX	CC	XX	Db 234 LysGluAsnAlaGlyAlaAlaGluLysAlaIleSerValHisSerThrProGluGlyCys 253
XX	CC	XX	QY 395 TCCTCCGCTTCTAGATGATCTTCGAGATTATGCATAAAGAGCTAAGGACCAACAAACG 454
XX	CC	XX	Db 254 SerSerAlaCysLysMetIleGluGluIleValHisLysGluAlaLysAspThrLysThr 273
XX	CC	XX	QY 455 GGTGACGAGGTTCCTTCGAGATCCTCGCCCAATAAATCTTGTAGGGCGCTCTCATGGC 514
XX	CC	XX	Db 274 AlaAspGluValProLeuLysIleLeuAlaHisAsnAsnPheValGlyArgLeuIleGly 293
XX	CC	XX	QY 515 AAGGAGGACGGAACCTTGAGAGAGTACGAGACCATCATCTGTGAAGGGGGGCATC 634
XX	CC	XX	Db 294 LysGluGlyArgAsnLeuLysLysValGluGlnAspThrGluThrLysIleThrIleSer 313
XX	CC	XX	QY 575 TCGTTGCAAGACCTTACCTTTTACACCCCTGAGAGACCATCATCTGTGAAGGGGGGCATC 634
XX	CC	XX	Db 314 SerLeuGlnAspLeuThrLeuTyrAsnProGluArgThrIleThrValLysGlyAlaIle 333
XX	CC	XX	QY 635 GAGAAATCTTCGAGGGCGGAGCAGGAAATATGAGAAAGTTCGGAGGCGCTATGAGAT 694
XX	CC	XX	Db 334 GluAsnCysCysArgAlaGluGlnGluIleMetLysLysValArgGluAlaTyrGluAsn 353
XX	CC	XX	QY 695 GATGTGGCTGCCATGAGC 712
XX	CC	XX	US-09-270-437D-5 (1-1708) x AAY30649 (1-577)
XX	CC	XX	QY 215 GTGGACATCCCTTCGGCTCGTCCGCCACCCAGTATGTGGTGCCATTATGGCAAG 274
XX	CC	XX	Db 194 ValAspIleProLeuArgLeuValProThrGlnTyrValGlyAlaIleGlyLys 213
XX	CC	XX	QY 275 GAGGGGGCCACATCCGCAACATCACAACACAGACCCAGTCCAAAGATAGACGTGCATAGG 334
XX	CC	XX	Db 214 GluGlyAlaThrIleArgAsnIleThrLysGlnThrGlnSerLysIleAspValHisArg 233
XX	CC	XX	QY 335 AAGGAGAACGGAGTGCAGCTGAGTGAAGAACCATCATGTGTGCACCTCCACCCCTGAGGGCTGC 394
XX	CC		

Db 354 AspValAlaAlaMetSer 359
RESULT 5
AAU16579
ID AAU16579 standard; protein; 250 AA.
XX AC AAU16579;
XX DT
XX DT 07-NOV-2001 (first entry)
XX DE Human novel secreted protein, Seq ID 1532.
XX KW Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic;
KW cardiant; vasotropic; cerebroprotective; nootropic; neuroprotective;
KW antibacterial; virucide; fungicide; opthalmological; vulnary;
KW secreted protein; rheumatoid arthritis; hyperproliferative disorder;
KW cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
KW cerebral ischaemia; angiogenesis; nervous system disorder;
KW Alzheimer's disease; infection; ocular disorder; corneal infection;
KW wound healing; epithelial cell proliferation; skin ageing; food additive;
KW preservative; antiproliferative.
XX KW
XX OS Homo sapiens.
XX FN WO200155322-A2.
XX PD
XX PD 02-AUG-2001.
XX PF
XX PF 17-JAN-2001; 2001WO-US001341.
XX PR
XX PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209457P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217497P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225466P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229257P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234977P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236388P.
PR 29-SEP-2000; 2000US-0236389P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 17-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.

PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249246P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 03-DEC-2000; 2000US-0251030P.
PR 03-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
PI XX
XX WPI; 2001-488783/53.
DR N-PSDB; AAS26566.
XX
XX New nucleic acid molecules encoding 461 human secreted proteins for
PT diagnosing, preventing, treating or ameliorating medical conditions and
PT used as food additives or preservatives.
XX
XX Claim 11; SEQ ID NO 1532; 980pp; English.
XX
XX The invention relates to isolated nucleic acid molecules and their
CC encoded secreted proteins. The nucleic acids and proteins are used to
CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
CC in diagnosing a pathological condition or susceptibility to a
CC pathological condition. Antibodies to the proteins can also be used in
CC alleviating symptoms associated with the disorders and in diagnostic
CC immunoassays e.g. radioimmunoassays or enzyme linked immunosorbant assays
CC (ELISA). Disorders which are diagnosed or treated include autoimmune
CC diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.
CC neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac
CC arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis,
CC nervous system disorders e.g. Alzheimer's disease, infections caused by
CC bacteria, viruses and fungi and ocular disorders e.g. corneal infection,
CC and many other disorders listed in the specification. The polypeptides
CC can also be used to aid wound healing and epithelial cell proliferation,
CC to prevent skin aging due to sunburn, to maintain organs before
CC transplantation, for supporting cell culture of primary tissues, to
CC regenerate tissues and in chemotaxis. The polypeptides can also be used
CC as a food additive or preservative to increase or decrease storage
CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,
CC minerals, cofactors and other nutritional components. The present
CC sequence represents a novel secreted protein of the invention. Note: The
CC sequence data for this patent did not form part of the printed

PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249246P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 03-DEC-2000; 2000US-0251030P.
PR 03-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
PI XX
XX WPI; 2001-488783/53.
DR N-PSDB; AAS26566.
XX
XX New nucleic acid molecules encoding 461 human secreted proteins for
PT diagnosing, preventing, treating or ameliorating medical conditions and
PT used as food additives or preservatives.
XX
XX Claim 11; SEQ ID NO 1532; 980pp; English.
XX
XX The invention relates to isolated nucleic acid molecules and their
CC encoded secreted proteins. The nucleic acids and proteins are used to
CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
CC in diagnosing a pathological condition or susceptibility to a
CC pathological condition. Antibodies to the proteins can also be used in
CC alleviating symptoms associated with the disorders and in diagnostic
CC immunoassays e.g. radioimmunoassays or enzyme linked immunosorbant assays
CC (ELISA). Disorders which are diagnosed or treated include autoimmune
CC diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.
CC neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac
CC arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis,
CC nervous system disorders e.g. Alzheimer's disease, infections caused by
CC bacteria, viruses and fungi and ocular disorders e.g. corneal infection,
CC and many other disorders listed in the specification. The polypeptides
CC can also be used to aid wound healing and epithelial cell proliferation,
CC to prevent skin aging due to sunburn, to maintain organs before
CC transplantation, for supporting cell culture of primary tissues, to
CC regenerate tissues and in chemotaxis. The polypeptides can also be used
CC as a food additive or preservative to increase or decrease storage
CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,
CC minerals, cofactors and other nutritional components. The present
CC sequence represents a novel secreted protein of the invention. Note: The
CC sequence data for this patent did not form part of the printed

Alignment Scores:
Pred. No.: 4,12e-105 Length: 250
Score: 116.00 Matches: 116
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 20.71% Indels: 0
DB: 4 Gaps: 0

US-09-270-437D-5 (1-1708) x AAU16579 (1-250)

QY 1013 GCCCAATTCAAGCTCAGGAGAGATCTATGGCAACTCAAGAGAGAGACTCTTTGGT 1072
DB 135 AlaGlnPheLysAlaGlnGlyArgIleGlyLysLeuLysGluAsnPheGly 154

QY 1073 CCCAAGGAGGAAAGTGAAGCTGGAGACCCACATACGTGTGTCAGCATACGAGCTGGCCGG 1132
DB 155 ProLysGluValLysLeuGluThrHisIleArgValProAlaSerAlaAlaGlyArg 174
QY 1133 GTCATTGGCAAGGTGGAAAAACCGTGAACCGATTGCAGATTTCACGCGCAGCTGAGGTG 1192
DB 175 VallileGlyLysGlyGlyThrValAsnGluLeuGlnAsnLeuThrAlaAlaGluVal 194
QY 1193 GTAGTACCAAGAGACACGACCCCTGATGAGAACACGAGTCACTCGTGAATAATCATCGGA 1252
DB 195 ValValProArgAspGlnThrProAspGluAsnAspGlnValIleValLysIleIleGly 214
QY 1253 CATTTCTATGTCAGTCAAGTGGCTCAAGGGAAGATCCGAGACATCCTGGCCCGAGTTAAG 1312
DB 215 HisPheTyrAlaSerGlnMetAlaGlnArgLysIleArgAspIleLeuAlaGlnValLys 234
QY 1313 CAGCAGCATCAGAAGGACAGAGTAACACGAGCCCGCAGGACGAGGAAG 1360
DB 235 GlnGlnHisGlnLysGlyGlnSerAsnGlnAlaGlnAlaArgGlyLys 250
RESULT 6
ABU55648
ID ABU55648 standard; protein; 250 AA.
XX
AC ABU55648;
XX
DT 18-MAR-2003 (first entry)
XX
DE Human novel polypeptide #735.
XX
KW Human; neural disorder; immune system disorder; renal disorder;
KW muscular disorder; respiratory disease; reproductive disorder;
KW gastrointestinal disorder; pulmonary disorder; cardiovascular disorder;
KW hyperproliferative disorder; inflammatory disease; allergic reaction;
KW blood related disorder; cancer; immunosuppressive; antiinflammatory;
KW cardiovascular; nephrotropic; cytostatic; antiallergic; thrombolytic;
KW haemostatic; antiarteriosclerotic.
XX
OS Homo sapiens.
XX
PN US2002132753-A1.
XX
PD 19-SEP-2002.
XX
XX 17-JAN-2001; 2001US-00764864.
XX
XX 31-JAN-2000; 2000US-0179065P.
XX 04-FEB-2000; 2000US-0180628P.
XX 28-JUN-2000; 2000US-0214886P.
XX 07-JUL-2000; 2000US-0216647P.
XX 07-JUL-2000; 2000US-0216880P.
XX 11-JUL-2000; 2000US-0217487P.
XX 11-JUL-2000; 2000US-0217496P.
XX 14-JUL-2000; 2000US-0218290P.
XX 26-JUL-2000; 2000US-0220963P.
XX 26-JUL-2000; 2000US-0220964P.
XX 14-AUG-2000; 2000US-0224518P.
XX 14-AUG-2000; 2000US-0224519P.
XX 14-AUG-2000; 2000US-0225267P.
XX 14-AUG-2000; 2000US-0225268P.
XX 14-AUG-2000; 2000US-0225270P.
XX 14-AUG-2000; 2000US-0225447P.
XX 14-AUG-2000; 2000US-0225757P.
XX 14-AUG-2000; 2000US-0225758P.
XX 22-AUG-2000; 2000US-0226868P.
XX 30-AUG-2000; 2000US-0228924P.
XX 01-SEP-2000; 2000US-0228928P.
XX 01-SEP-2000; 2000US-0229343P.
XX 01-SEP-2000; 2000US-0229344P.
XX 01-SEP-2000; 2000US-0229345P.
XX 05-SEP-2000; 2000US-0229509P.
XX 05-SEP-2000; 2000US-0229513P.
XX 08-SEP-2000; 2000US-0231413P.

```

PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 27-SEP-2000; 2000US-0235834P.
PR 28-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241809P.
PR 01-NOV-2000; 2000US-0244617P.
PR 17-NOV-2000; 2000US-0249299P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
XX
XX (ROSE/) ROSEN C A.
PA (RUBE/) RUBEN S M.
PA (BARA/) BARASH S C.
XX
XX Rosen CA, Ruben SM, Barash SC;
PI
XX WPI: 2003-147444/14.
DR N-PSDB; ABX73907.
XX
XX New polypeptides and nucleic acids, useful in gene therapy for treating,
PT inhibiting or preventing e.g. neural, immune system, muscular,
PT respiratory, reproductive, gastrointestinal, pulmonary, cardiovascular or
PT renal disorders.
XX
XX Claim 11; SEQ ID NO 1532; 402pp; English.
XX
XX The invention relates to human novel polypeptides and their associated
XX polynucleotides. The polypeptides and polynucleotides are useful in gene
XX therapy for treating, inhibiting or preventing neural disorders, immune
XX system disorders (e.g. systemic lupus erythematosus, rheumatoid arthritis
XX and multiple sclerosis), muscular disorders, respiratory diseases (e.g.
XX nasal vestibulitis, nasal polyps and sinusitis), reproductive disorders,
XX gastrointestinal disorders, pulmonary disorders, cardiovascular disorders,
XX (e.g. congenital heart defects, Ebstein's anomaly and hypoplastic left
XX heart syndrome), renal disorders (e.g. acute kidney failure and end-stage
XX renal disease), hyperproliferative disorders (e.g. Hodgkin's disease and
XX leukaemia), inflammatory diseases (e.g. septic shock, bursitis and
XX appendicitis), allergic reactions and conditions (e.g. asthma), blood
XX related disorders (e.g. thrombosis, atherosclerosis and myocardial
XX infarction) and cancerous diseases. Sequences ABU54914-ABU55699 and
XX ABU55748 represent human novel polypeptides of the invention
XX
XX Sequence 250 AA;
XX
XX Alignment Scores:
XX Pred. No.: 4,12e-105 Length: 250
XX Score: 116.00 Matches: 116
XX Percent Similarity: 100.00% Conservative: 0
XX Best Local Similarity: 100.00% Mismatches: 0
XX Query Match: 20.71% Indels: 0
XX DB: 6 Gaps: 0
XX
XX US-09-270-437D-5 (1-1708) x ABU55648 (1-250)
XX
XX 1013 GCCAATTCAGGCTCAGGGAAGATCTATGCGAAACTCAGGAGGAGAACTTCTTGGT 1072
XX
XX 135 AlaglnPhelysAlaGlnGlyArgIleTyGlyLysLeuLysGluGluAsnPheGly 154
XX
XX 1073 CCCAAGGAGGAAGTGAAGCTGGAGCCACCATACGTTGCCAGCATCAGCAGCTGCCGCG 1132

```

```

Db 155 ProLysGluGluValLysLeuGluThrHisIleArgValProAlaSerAlaAlaGlyArg 174
QY 1133 GTCATTGGCAAGGTGGAACACGGTGAACGAGTTGCAGAAATTTGACGGCAGCTGAGGTG 1192
Db 175 ValIleGlyLysGlyLysThrValAsnGluLeuGlnAsnLeuThrAlaAlaGluVal 194
QY 1193 GTAGTACCAAGACACGACCCCTGATGAGAACGACACGAGTCTATCGTAAATCATCGGA 1252
Db 195 ValValProArgAspGlnThrProAspGluAsnAspGlnValIleValIleGly 214
QY 1253 CATTTCTATGCCAGTCAGATGGCTCAACGGAAGTACCGAGCATCTGCCCCAGGTTAAG 1312
Db 215 HisPheTyRAlaSerGlnMetAlaGlnArgLysIleArgAspIleLeuAlaGlnVal 234
QY 1313 CAGCAGCATCAGAGGACAGAGTAACCGAGCCCGACGACGAGGAG 1360
Db 235 GlnGlnHisGlnLysGlyGlnSerAsnGlnAlaGlnAlaArgLys 250

RESULT 7
AAB11328
ID AAB11328 standard; protein; 579 AA.
XX
XX AAB11328;
XX
XX 21-FEB-2001 (first entry)
XX
XX Human lung cancer-associated protein L523S.
XX
XX Lung cancer; therapy; treatment; human; tumor; immunogenic; cytostatic;
XX vaccine; detection.
XX
XX Homo sapiens.
XX
XX WO2000061612-A2.
XX
XX 19-OCT-2000.
XX
XX 03-APR-2000; 2000WO-US0008896.
XX
XX 02-APR-1999; 99US-00285479.
XX 17-DEC-1999; 99US-00466396.
XX 30-DEC-1999; 99US-00476496.
XX 10-JAN-2000; 2000US-00480884.
XX 22-FEB-2000; 2000US-00510376.
XX
XX (CORI-) CORIXA CORP.
XX
XX Wang T, Fan L;
XX
XX WPI; 2000-628399/60.
XX N-PSDB; AAC65900.
XX
XX Isolated polypeptide comprising an immunogenic portion of a lung tumor
XX protein is used for detecting and monitoring progression of lung cancer
XX in a patient.
XX
XX Claim 3; Page 186-188; 261pp; English.
XX
XX This invention describes a novel isolated polypeptide (I) which
XX comprising an immunogenic portion of a lung tumor protein or variant (P2)
XX which have cytostatic activity. The polypeptides and polynucleotides are
XX used in compositions and vaccines to inhibit the development of cancer,
XX especially lung cancer, in a patient. Methods described in the invention
XX can be used to monitor the progression of a cancer by carrying out the
XX detection at subsequent time points and comparing the results from the
XX different time points. CD4+ and/or CD8+ T-Cells isolated from a patient
XX are treated with P2, polynucleotides encoding P2 or antigen presenting
XX cells expressing P2 and then administered to the patient to inhibit
XX development of cancer
XX
XX Sequence 579 AA;

```

Alignment Scores:

Pred. No.: 3,86e-26 Length: 579
 Score: 36.00 Matches: 36
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 6.43% Indels: 0
 DB: 3 Gaps: 0

US-09-270-437D-5 (1-1708) x AAB11328 (1-579)

QY 254 GTGGTCCCATTTGCGAAGAGGGGGCCACCATCCGCAACATCACAACAGACCCAG 313
 |||||
 Db 207 ValGlyAlaIleIleGlyLysGluGlyAlaThrIleArgAsnIleThrLysGlnThrGln 226
 |||||
 QY 314 TCCAAGATAGACGTGCATAGGAAGAGAACGCGAGGTGCAGCTGAAAAA 361
 |||||
 Db 227 SerLysIleAspValHisArgLysGluAsnAlaGlyAlaAlaGluLys 242
 |||||

RESULT 8

AAB11365
 ID AAB11365 standard; protein; 579 AA.
 XX AC AAB11365;
 XX DT 21-FEB-2001 (first entry)
 XX DE Human lung cancer associated antigen L523S.
 XX KW Lung cancer; therapy; treatment; human; tumor; immunogenic; cytostatic;
 XX KM vaccine; detection.
 XX OS Homo sapiens.
 XX PN WO200061612-A2.
 XX PD 19-OCT-2000.
 XX PF 03-APR-2000; 2000WO-US008896.
 XX PR 02-APR-1999; 99US-00285479.
 XX PR 17-DEC-1999; 99US-00466396.
 XX PR 30-DEC-1999; 99US-00476496.
 XX PR 10-JAN-2000; 2000US-00480884.
 XX PR 22-FEB-2000; 2000US-00510376.
 XX PA (CORI-) CORIXA CORP.
 XX PI Wang T, Fan L;
 XX WPI; 2000-628399/60.
 XX DR N-PSDB; AAC66035.
 XX PT Isolated polypeptide comprising an immunogenic portion of a lung tumor
 PT protein is used for detecting and monitoring progression of lung cancer
 PT in a patient.
 XX PS Claim 3; Page 259-261; 261pp; English.
 XX CC This invention describes a novel isolated polypeptide (I) which
 CC comprising an immunogenic portion of a lung tumor protein or variant (P2)
 CC which have cytostatic activity. The polypeptides and polynucleotides are
 CC used in compositions and vaccines to inhibit the development of cancer,
 CC especially lung cancer, in a patient. Methods described in the invention
 CC can be used to monitor the progression of a cancer by carrying out the
 CC detection at subsequent time points and comparing the results from the
 CC different time points. CD4+ and/or CD8+ T-cells isolated from a patient
 CC are treated with P2, polynucleotides encoding P2 or antigen presenting
 CC cells expressing P2 and then administered to the patient to inhibit
 CC development of cancer
 XX SQ Sequence 579 AA;

Alignment Scores:

Pred. No.: 3,86e-26 Length: 579
 Score: 36.00 Matches: 36
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 6.43% Indels: 0
 DB: 3 Gaps: 0

US-09-270-437D-5 (1-1708) x AAB11365 (1-579)

QY 254 GTGGTCCCATTTGCGAAGAGGGGGCCACCATCCGCAACATCACAACAGACCCAG 313
 |||||
 Db 207 ValGlyAlaIleIleGlyLysGluGlyAlaThrIleArgAsnIleThrLysGlnThrGln 226
 |||||
 QY 314 TCCAAGATAGACGTGCATAGGAAGAGAACGCGAGGTGCAGCTGAAAAA 361
 |||||
 Db 227 SerLysIleAspValHisArgLysGluAsnAlaGlyAlaAlaGluLys 242
 |||||

RESULT 9

ABE74960
 ID ABB74960 standard; protein; 579 AA.
 XX AC ABB74960;
 XX DT 01-MAY-2002 (first entry)
 XX DE Human lung tumour L523S protein sequence SEQ ID NO:176.
 XX KW Human; lung tumour; lung cancer; cytostatic; immunostimulant; vaccine;
 XX KM immune response.
 XX OS Homo sapiens.
 XX PN WO200200174-A2.
 XX PD 03-JAN-2002.
 XX PF 28-JUN-2001; 2001WO-US021065.
 XX PR 28-JUN-2000; 2000US-00606421.
 XX PR 02-AUG-2000; 2000US-00630940.
 XX PR 21-AUG-2000; 2000US-00643597.
 XX PR 15-SEP-2000; 2000US-00662786.
 XX PR 09-OCT-2000; 2000US-00685696.
 XX PR 12-DEC-2000; 2000US-00735705.
 XX PR 07-MAY-2001; 2001US-00850716.
 XX PA (CORI-) CORIXA CORP.
 XX PI Wang T, Wang A, Skeiky YAW, Li SX, Kalos MD, Henderson RA;
 XX PI McNeill PD, Fanger N, Retter MW, Marnierakis M, Fanger GR,
 XX PI Vedvick TS, Carter D, Watanabe Y, Peckham DW;
 XX WPI; 2002-090513/12.
 XX DR N-PSDB; ABL49119.
 XX PT Polynucleotides encoding lung tumor polypeptides, useful for treating
 PT lung cancer or stimulating an immune response.
 XX PS Example 2; Page 267-268; 374pp; English.
 XX CC The present invention describes human lung tumour proteins. Human lung
 CC tumour proteins and polynucleotides have cytostatic and immunostimulant
 CC activities, and can be used in vaccine production. Compositions
 CC comprising the lung tumour proteins, polynucleotides, antibodies, fusion
 CC proteins, T cell populations, or antigen presenting cells that express
 CC the lung tumour proteins are useful for treating lung cancer or
 CC stimulating an immune response. ABL48959 to ABL49300 and ABB74946 to
 CC ABB75070 represent sequences used in the exemplification of the present
 CC invention
 XX SQ Sequence 579 AA;

Alignment Scores:

Pred. No.: 3.86e-26 Length: 579
 Score: 36.00 Matches: 36
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 6.43% Indels: 0
 DB: 5 Gaps: 0

US-09-270-437D-5 (1-1708) x ABB74960 (1-579)

QY 254 GTGGTGCCATTATTGCAAGAGGGGGCCACCATCCGCAACATCACAACAGACCCAG 313
 Db 207 ValGlyAlaIleIleGlyGlyGluGlyAlaThrIleArgAsnIleThrGln 226
 QY 314 TCCAAGATAGACGTGCATAGGAGGAGACGCGAGGTGCAGCTGAAAAA 361
 Db 227 SerLysIleAspValHisArgLysGluAsnAlaGlyAlaAlaGluLys 242

RESULT 10

ABB75053
 ID ABB75053 standard; protein; 579 AA.

AC ABB75053;

XX 01-MAY-2002 (first entry)

DT Human lung tumour L523S recombinant protein sequence SEQ ID NO:446.

DE Human; lung tumour; lung cancer; cytostatic; immunostimulant; vaccine;
 KW immune response.

XX Homo sapiens.

XX WO200200174-A2.

XX 03-JAN-2002.

XX 28-JUN-2001; 2001WO-US021065.

XX 28-JUN-2000; 2000US-00606421.

XX 02-AUG-2000; 2000US-00630940.

XX 21-AUG-2000; 2000US-00643597.

XX 15-SEP-2000; 2000US-00662786.

XX 09-OCT-2000; 2000US-00685696.

XX 12-DEC-2000; 2000US-00735705.

XX 07-MAY-2001; 2001US-00850716.

XX (CORI-) CORIXA CORP.

XX Wang T, Wang A, Skeiky YAW, Li SX, Kalos MD, Henderson RA;

XX McNeill PD, Fanger N, Retter MW, Marnerakis M, Fanger GR;

XX Vedvick TS, Carter D, Watanabe Y, Peckham DW;

XX WPI; 2002-090513/12.

XX N-PSDB; ABL49297.

XX Polynucleotides encoding lung tumor polypeptides, useful for treating

XX lung cancer or stimulating an immune response.

XX Claim 2; Page 365-367; 374pp; English.

XX The present invention describes human lung tumour proteins. Human lung
 CC tumour proteins and polynucleotides have cytostatic and immunostimulant
 CC activities, and can be used in vaccine production. Compositions
 CC comprising the lung tumour proteins, polynucleotides, antibodies, fusion
 CC proteins, T cell populations, or antigen presenting cells that express
 CC the lung tumour proteins are useful for treating lung cancer or
 CC stimulating an immune response. ABL48959 to ABL49300 and ABB74946 to
 CC ABB75070 represent sequences used in the exemplification of the present
 CC invention

XX Sequence 579 AA;

XX Alignment Scores:

Pred. No.: 3.86e-26 Length: 579
 Score: 36.00 Matches: 36
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 6.43% Indels: 0
 DB: 5 Gaps: 0

US-09-270-437D-5 (1-1708) x ABB75053 (1-579)

QY 254 GTGGTGCCATTATTGCAAGAGGGGGCCACCATCCGCAACATCACAACAGACCCAG 313
 Db 207 ValGlyAlaIleIleGlyGlyGluGlyAlaThrIleArgAsnIleThrGln 226
 QY 314 TCCAAGATAGACGTGCATAGGAGGAGACGCGAGGTGCAGCTGAAAAA 361
 Db 227 SerLysIleAspValHisArgLysGluAsnAlaGlyAlaAlaGluLys 242

RESULT 11

ABB74997

ID ABB74997 standard; protein; 579 AA.

AC ABB74997;

XX 01-MAY-2002 (first entry)

DT Human lung tumour L523S protein sequence SEQ ID NO:348.

DE Human; lung tumour; lung cancer; cytostatic; immunostimulant; vaccine;
 KW immune response.

XX Homo sapiens.

XX WO200200174-A2.

XX 03-JAN-2002.

XX 28-JUN-2001; 2001WO-US021065.

XX 28-JUN-2000; 2000US-00606421.

XX 02-AUG-2000; 2000US-00630940.

XX 21-AUG-2000; 2000US-00643597.

XX 15-SEP-2000; 2000US-00662786.

XX 09-OCT-2000; 2000US-00685696.

XX 12-DEC-2000; 2000US-00735705.

XX 07-MAY-2001; 2001US-00850716.

XX (CORI-) CORIXA CORP.

XX Wang T, Wang A, Skeiky YAW, Li SX, Kalos MD, Henderson RA;

XX McNeill PD, Fanger N, Retter MW, Marnerakis M, Fanger GR;

XX Vedvick TS, Carter D, Watanabe Y, Peckham DW;

XX WPI; 2002-090513/12.

XX N-PSDB; ABL49254.

XX Polynucleotides encoding lung tumor polypeptides, useful for treating

XX lung cancer or stimulating an immune response.

XX Example 2; Page 330-332; 374pp; English.

XX The present invention describes human lung tumour proteins. Human lung
 CC tumour proteins and polynucleotides have cytostatic and immunostimulant
 CC activities, and can be used in vaccine production. Compositions
 CC comprising the lung tumour proteins, polynucleotides, antibodies, fusion
 CC proteins, T cell populations, or antigen presenting cells that express
 CC the lung tumour proteins are useful for treating lung cancer or
 CC stimulating an immune response. ABL48959 to ABL49300 and ABB74946 to
 CC ABB75070 represent sequences used in the exemplification of the present
 CC invention

XX Sequence 579 AA;

XX Alignment Scores:

Alignment Scores:

SQ Sequence 579 AA;
 Alignment Scores:
 Pred. No.: 3,86e-26 Length: 579
 Score: 36.00 Matches: 36
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 6.43% Indels: 0
 DB: 5 Gaps: 0

US-09-270-437D-5 (1-1708) x ABP61917 (1-579)

Qy 254 GTGGGTGCATTATTCGCAAGGAGGGCCACCATCGCAACATCAAAACAGACCCAG 313
 Db 207 ValGlyAlaIleIleGlyLysGluGlyAlaThrIleArgAsnIleThrLysGlnThrGln 226
 Qy 314 TCCAGATAGACGTGCATAGGAGGAGAGCGACGTCAGCTGAAAAA 361
 Db 227 SerLysIleAspValHisArgLysGluAsnAlaGlyAlaAlaGluLys 242

RESULT 14
 ABP61974
 ID ABP61974 standard; protein; 579 AA.
 AC ABP61974;
 XX
 DT 07-OCT-2002 (first entry)
 DE Human lung cancer associated protein sequence SEQ ID NO:449.
 KW Human; lung cancer; lung tumour; cytostatic; gene therapy; vaccine.
 XX Homo sapiens.
 OS
 FN WO200247534-A2.
 PD 20-JUN-2002.
 XX
 PF 30-NOV-2001; 2001WO-US047576.
 PR 12-DEC-2000; 2000US-00735705.
 PR 07-MAY-2001; 2001US-00850716.
 PR 28-JUN-2001; 2001US-00897778.
 XX
 PA (CORI-) CORIXA CORP.
 PI Wang T, Wang A, Skeiky YAW, Li SX, Kalos MD, Henderson RA;
 PI McNeill PD, Fanger N, Retter MW, Durham M, Fanger GR, Vedvick TS;
 PI Carter D, Watanabe Y, Peckham DW, Cai F, Foy TM;
 DR WPI; 2002-583465/62.
 DR N-PSDB; ABQ92485.
 XX
 PT Novel lung carcinoma polynucleotide sequences and polypeptides encoded by
 PT the polynucleotides, useful in pharmaceutical compositions such as
 PT vaccines and as markers to indicate the presence of lung cancer.
 XX
 PS Claim 9; Page 375-377; 381pp; English.
 XX
 CC The present invention describes isolated human lung carcinoma
 CC polynucleotides (I) and polypeptides (II). (I) and (II) have cytostatic
 CC activity, and can be used in gene therapy and in vaccines. Compositions
 CC comprising (I) or (II) can be used for stimulating an immune response in
 CC a patient and for treating lung cancer in a patient. Oligonucleotides of
 CC (I) can be used for detecting the presence of a cancer in a patient, by
 CC obtaining a biological sample from the patient, contacting the biological
 CC sample with the oligonucleotide, detecting in the sample, an amount of
 CC polynucleotide that hybridises to the oligonucleotide and comparing the
 CC amount of polynucleotide that hybridises to the oligonucleotide to a
 CC predetermined cut-off value, and determining the presence of a cancer in
 CC the patient. (I) and (II) are useful in pharmaceutical compositions, e.g.
 CC vaccines. (I) is useful as a marker to indicate the presence or absence
 CC of a cancer such as lung cancer. ABQ92145 to ABQ92486 and ABP61866 to

CC ABP61992 represent sequences used in the exemplification of the present
 CC invention.
 XX SQ Sequence 579 AA;
 Alignment Scores:
 Pred. No.: 3,86e-26 Length: 579
 Score: 36.00 Matches: 36
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 6.43% Indels: 0
 DB: 5 Gaps: 0

US-09-270-437D-5 (1-1708) x ABP61974 (1-579)

Qy 254 GTGGGTGCATTATTCGCAAGGAGGGCCACCATCGCAACATCAAAACAGACCCAG 313
 Db 207 ValGlyAlaIleIleGlyLysGluGlyAlaThrIleArgAsnIleThrLysGlnThrGln 226
 Qy 314 TCCAGATAGACGTGCATAGGAGGAGAGCGACGTCAGCTGAAAAA 361
 Db 227 SerLysIleAspValHisArgLysGluAsnAlaGlyAlaAlaGluLys 242

RESULT 15
 ABP61880
 ID ABP61880 standard; protein; 579 AA.
 AC ABP61880;
 XX
 DT 07-OCT-2002 (first entry)
 DE Human lung cancer associated protein sequence SEQ ID NO:176.
 KW Human; lung cancer; lung tumour; cytostatic; gene therapy; vaccine.
 XX Homo sapiens.
 OS
 FN WO200247534-A2.
 PD 20-JUN-2002.
 XX
 PF 30-NOV-2001; 2001WO-US047576.
 PR 12-DEC-2000; 2000US-00735705.
 PR 07-MAY-2001; 2001US-00850716.
 PR 28-JUN-2001; 2001US-00897778.
 XX
 PA (CORI-) CORIXA CORP.
 PI Wang T, Wang A, Skeiky YAW, Li SX, Kalos MD, Henderson RA;
 PI McNeill PD, Fanger N, Retter MW, Durham M, Fanger GR, Vedvick TS;
 PI Carter D, Watanabe Y, Peckham DW, Cai F, Foy TM;
 DR WPI; 2002-583465/62.
 DR N-PSDB; ABQ92305.
 XX
 PT Novel lung carcinoma polynucleotide sequences and polypeptides encoded by
 PT the polynucleotides, useful in pharmaceutical compositions such as
 PT vaccines and as markers to indicate the presence of lung cancer.
 XX
 PS Example 2; Page 274-275; 381pp; English.
 XX
 CC The present invention describes isolated human lung carcinoma
 CC polynucleotides (I) and polypeptides (II). (I) and (II) have cytostatic
 CC activity, and can be used in gene therapy and in vaccines. Compositions
 CC comprising (I) or (II) can be used for stimulating an immune response in
 CC a patient and for treating lung cancer in a patient. Oligonucleotides of
 CC (I) can be used for detecting the presence of a cancer in a patient, by
 CC obtaining a biological sample from the patient, contacting the biological
 CC sample with the oligonucleotide, detecting in the sample, an amount of
 CC polynucleotide that hybridises to the oligonucleotide and comparing the
 CC amount of polynucleotide that hybridises to the oligonucleotide to a
 CC predetermined cut-off value, and determining the presence of a cancer in
 CC the patient. (I) and (II) are useful in pharmaceutical compositions, e.g.
 CC vaccines. (I) is useful as a marker to indicate the presence or absence
 CC of a cancer such as lung cancer. ABQ92145 to ABQ92486 and ABP61866 to